

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2004, 21:00:09 ; Search time 1434 Seconds
(without alignments)
7322.251 Million cell updates/sec

Title: US-09-993-966-5

Perfect score: 1859

Sequence: 1 gaattcgcccttctaatacg.....aaaaaaaaaagcgccgc 1859

Scoring table: ORIGO_NUC Gapop 60.0 , Capext 60.0

Searched: 416236 seqs, 2824127955 residue

Word size : 0

Total number of hits satisfying chosen parameters: 8352472

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*

16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*

17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*

18: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*

19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*

20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1859	100.0	1859	10	US-09-993-966-5 Sequence 5, Appl1
2	1786	96.1	1786	10	US-09-993-966-1 Sequence 1, Appl1
3	1417	76.2	1417	13	US-10-087-192-179 Sequence 179, App
4	783	42.1	2379	10	US-09-764-891-5902 Sequence 5902, Ap
5	783	42.1	2379	10	US-09-764-891-5904 Sequence 5904, Ap
6	783	42.1	106315	13	US-10-087-192-178 Sequence 178, App
7	590	31.7	590	10	US-09-993-966-21 Sequence 21, App
8	377	20.3	427	10	US-09-918-995-8686 Sequence 8686, Ap
9	215	11.6	215	10	US-09-993-966-22 Sequence 22, Appl
10	148	8.0	148	10	US-09-993-966-18 Sequence 18, Appl
11	137	7.4	598	15	US-10-029-386-7990 Sequence 7990, Ap
12	137	7.4	1950	13	US-10-016-157A-126 Sequence 126, App

13	134	7.2	134	10	US-09-993-966-14 Sequence 14, Appl
14	133	7.2	133	15	US-10-029-386-21690 Sequence 21690, A
15	128	6.9	128	10	US-09-993-966-20 Sequence 20, Appl
16	120	6.5	120	10	US-09-993-966-10 Sequence 10, Appl
17	108	5.8	639	15	US-10-243-552-231 Sequence 231, App
18	107	5.8	107	10	US-09-993-966-16 Sequence 16, Appl
19	96	5.2	96	10	US-09-993-966-17 Sequence 17, Appl
20	88	4.7	554	15	US-10-029-386-7615 Sequence 7615, Ap
21	86	4.6	179	15	US-10-029-386-21315 Sequence 21315, A
22	85	4.6	85	10	US-09-993-966-19 Sequence 19, Appl
23	68	3.7	566	13	US-10-027-632-266931 Sequence 266931, Sequence 266931, Sequence 1214, Ap
24	68	3.7	566	15	US-10-027-632-266931 Sequence 1214, Ap
25	68	3.7	2724	9	US-09-833-381-1215 Sequence 1215, Ap
26	68	3.7	2724	9	US-09-833-381-1215 Sequence 1215, Ap
27	67	3.6	67	10	US-09-993-966-15 Sequence 15, Appl
28	58	3.1	670	10	US-10-275-287-19 Sequence 19, Appl
29	57	3.1	417	16	US-10-275-287-28 Sequence 28, Appl
30	57	3.1	510	16	US-10-275-287-17 Sequence 17, Appl
31	56	3.0	636	16	US-10-275-287-25 Sequence 25, Appl
32	56	3.0	651	16	US-10-275-287-8 Sequence 8, Appl1
33	56	3.0	666	16	US-10-275-287-18 Sequence 18, Appl
34	56	3.0	707	15	US-10-072-602B-214 Sequence 214, App
35	55	3.0	547	16	US-10-275-287-10 Sequence 10, Appl
36	55	3.0	1245	16	US-10-275-287-5 Sequence 5, Appl1
37	55	3.0	1245	16	US-10-275-287-52 Sequence 52, Appl
38	54	2.9	1874	17	US-10-635-223-16 Sequence 16, Appl
39	54	2.9	1874	17	US-10-635-223-16 Sequence 20, Appl
40	54	2.9	2605	17	US-10-635-223-42 Sequence 42, Appl
41	53	2.9	787	16	US-10-275-287-54 Sequence 54, Appl
42	53	2.9	1434	15	US-10-210-760A-27 Sequence 27, Appl
43	53	2.9	1434	16	US-10-229-541A-122 Sequence 122, App
44	53	2.9	3281	16	US-10-171-319-13 Sequence 13, Appl
45	53	2.9	3768	15	US-10-101-464A-949 Sequence 949, App

ALIGNMENTS

RESULT 1

US-09-993-966-5

Sequence 5, Application US/0993966

Publication No. US20030186232A1

GENERAL INFORMATION:

APPLICANT: ROMAN, MICHAEL

TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF MKD PROTEIN,

FILE REFERENCE: 014024/0280733

CURRENT APPLICATION NUMBER: US/09/993, 966

CURRENT FILING DATE: 2001-11-27

PRIOR APPLICATION NUMBER: 60/252, 884

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/291, 109

PRIOR FILING DATE: 2001-05-16

PRIOR APPLICATION NUMBER: 60/325, 571

PRIOR FILING DATE: 2001-10-01

NUMBER OF SEQ ID NOS: 26

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 5

LENGTH: 1859

TYPE: DNA

ORGANISM: Homo sapiens

US-09-993-966-5

Query Match 100.0%; Score 1859; DB 10; Length 1859;

Best local Similarity 100.0%; Pred. No. 0;

Matches 1859; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTGCGCTTTTAATAGACTCACTATAGGCAAGAGTGTAAACAGCAGACTAGC 60

1 GAATTGCGCTTTTAATAGACTCACTATAGGCAAGAGTGTAAACAGCAGACTAGC 60

CGGGAGTGGGGCGGGGCGGAGCGGCGGAGCGGCTCCGGCGGCTGGGCTCCG 120

61 CGGGAGTGGGCGGGCGAGCGGCGAGAGGCGGTCCCGGCGCCGCTCGGAGCTCCG 120
QY 121 CTGGGCTCGGGGGCTGTTTGGGAGAGAGAGGCGCAAGGGAGGCGCCAGGCGCGGGGCC 180
Db 121 CTGGGCTCGGGGGCTGTTTGGGAGAGAGAGGCGCAAGGGAGGCGCGCGGGGCC 180
QY 181 GGGCGCATGGCTTAGGGAGCGCTCCGGGCGCGCGAGCGCCAGAGTGGGAGAACTTCACTC 240
Db 181 GGGCGCATGGCTTAGGGAGCGCTCCGGGCGCGCGAGCGCCAGAGTGGGAGAACTTCACTC 240
QY 241 CAAGCGGGCGCGGTGTGCAAGCGCAAGAGAGCGCCGAGAGTGCAGCTTCGCGGTGAG 300
Db 241 CAAGCGGGCGCGGTGTGCAAGCGCAAGAGAGAGCGCCGAGAGTGCAGCTTCGCGGTGAG 300
QY 301 CGCTGCTGGGCTCGGAGAGGCGCATCGAGAGTGGATGGGAGACAGGCGTGGCCGGGCGG 360
Db 301 CGCTGCTGGGCTCGGAGAGGCGCATCGAGAGTGGATGGGAGACAGGCGTGGCCGGGCGG 360
QY 361 TGTCTCGGAGACCGCGACAGCTGCGGTTGGCGGGCACCATATGAGCCGAGCACCGGGAGCT 420
Db 361 TGTCTCGGAGACCGCGACAGCTGCGGTTGGCGGGCACCATATGAGCCGAGCACCGGGAGCT 420
QY 421 CGTGGCGCATGTTTGAAGACAAGCTCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Db 421 CGTGGCGCATGTTTGAAGACAAGCTCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 481 AGTGGCGCTGCTCTCTGAG 540
Db 481 AGTGGCGCTGCTCTCTGAG 540
QY 541 GAGAGTGAAG 600
Db 541 GAGAGTGAAG 600
QY 601 CGAGGTGTCAATGAG 660
Db 601 CGAGGTGTCAATGAG 660
QY 661 CAACAAAGGAG 720
Db 661 CAACAAAGGAG 720
QY 721 GGTGGAATCTCTGTCAACCACTCCCAACATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Db 721 GGTGGAATCTCTGTCAACCACTCCCAACATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 781 CGTGGCGCGCGAG 840
Db 781 CGTGGCGCGCGAG 840
QY 841 GAGGCGAAGGCGCGAG 900
Db 841 GAGGCGAAGGCGCGAG 900
QY 901 GCGAGAGAGCGCGCTCAAGTTTCCAGGAGTGAAGCGCGCTGAGAGAGAGAGAGAGAGAGAG 960
Db 901 GCGAGAGAGCGCGCTCAAGTTTCCAGGAGTGAAGCGCGCTGAGAGAGAGAGAGAGAGAGAG 960
QY 961 CCATTGGGTAGATAG 1020
Db 961 CCATTGGGTAGATAG 1020
QY 1021 AAATTAACAGCTCCCAATTTGGGCGTGGCTCCCTTCCGTGGCGCGAGAGTGAAGAGTGGC 1080
Db 1021 AAATTAACAGCTCCCAATTTGGGCGTGGCTCCCTTCCGTGGCGCGAGAGTGAAGAGTGGC 1080
QY 1081 CCCCAGACCTCAATCCCATCGATCGCTCCCATGAGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Db 1081 CCCCAGACCTCAATCCCATCGATCGCTCCCATGAGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1140
QY 1141 ACACCGAAGAGCGCGAG 1200
Db 1141 ACACCGAAGAGCGCGAG 1200

QY 1201 CAAGGCTCAGAGCTCAGCAAGGCTCGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Db 1201 CAAGGCTCAGAGCTCAGCAAGGCTCGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1261 GAGTCCCGGAG 1320
Db 1261 GAGTCCCGGAG 1320
QY 1321 AAACAAAGCGCGCTCTGAG 1380
Db 1321 AAACAAAGCGCGCTCTGAG 1380
QY 1381 CAGCGCGCGCT 1440
Db 1381 CAGCGCGCGCT 1440
QY 1441 CAAG 1500
Db 1441 CAAG 1500
QY 1501 CCGGGGCGGAG 1560
Db 1501 CCGGGGCGGAG 1560
QY 1561 GAGCGGCTCAG 1620
Db 1561 GAGCGGCTCAG 1620
QY 1621 CTTCTACAG 1680
Db 1621 CTTCTACAG 1680
QY 1681 CCGGACAG 1740
Db 1681 CCGGACAG 1740
QY 1741 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1800
Db 1741 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1800
QY 1801 CACAG 1859
Db 1801 CACAG 1859

RESULT 2
US-09-993-966-1
; Sequence 1, Application US/09993966
; Publication No. US20030186232A1
; GENERAL INFORMATION:
; APPLICANT: ROHAN, MICHAEL
; TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NMD PROTEIN,
; FILE REFERENCE: 014024/028073
; CURRENT APPLICATION NUMBER: US/09/993,966
; PRIOR FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 60/252,884
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/291,109
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/325,571
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1786
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-993-966-1

Query Match 96.1%; Score 1786; DB 10; Length 1786;
Best Local Similarity 100.0%; Pred. No. 0;

Query Match 76.2%; Score 1417; DB 13; Length 1448;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 CCCAGCATGGGAACTTCACTCCAGCCGCGCTGTGCAAGCGCAGGAGAGCCC 276
Db 32 CCCAGCATGGGAACTTCACTCCAGCCGCGCTGTGCAAGCGCAGGAGAGCCC 91
QY 277 GGAAGTGAACGTTCCGCGGTGAGCGCTGCTGGGCTCGAAGGGCATTCGAGAGTGGAT 336
Db 92 GGAAGTGAACGTTCCGCGGTGAGCGCTGCTGGGCTCGAAGGGCATTCGAGAGTGGAT 151
QY 337 CGGAGACAGCGCTGCGCGCGGTGTCTGTGGGACCCCGACAGCTGGGTGGGGGCGAC 396
Db 152 CGGAGACAGCGCTGCGCGCGGTGTCTGTGGGACCCCGACAGCTGGGTGGGGGCGAC 211
QY 397 CATAGCCGAAGCACCAGGAGCTCGTGGGCGACGTGTGAGAGACAGCTCAGCGAGGA 456
Db 212 CATAGCCGAAGCACCAGGAGCTCGTGGGCGACGTGTGAGAGACAGCTCAGCGAGGA 271
QY 457 AGAGAGAGACGACTTTCGGCTGGAAGTGGCCCTGCTTGAGAAACTGAAGGGCTGGG 516
Db 272 AGAGAGAGACGACTTTCGGCTGGAAGTGGCCCTGCTTGAGAAACTGAAGGGCTGGG 331
QY 517 CAGCGGAGATGAGAAAGAAAGATGAGAGAGGGAACCTGGCCAGGCTCCAGAAAGCA 576
Db 332 CAGCGGAGATGAGAAAGAAAGATGAGAGAGGGAACCTGGCCAGGCTCCAGAAAGCA 391
QY 577 GCTGAAGTTTGAAGAGCTCCAGTCCGACGTGTCCATGAGAGAGACAGCCGCGAGAGTG 636
Db 392 GCTGAAGTTTGAAGAGCTCCAGTCCGACGTGTCCATGAGAGAGACAGCCGCGAGAGTG 451
QY 637 GACCTTCAACCTGTATGACTTTTGACAACAGGGCAAGGTCAACCGAAGGACATCACCG 696
Db 452 GACCTTCAACCTGTATGACTTTTGACAACAGGGCAAGGTCAACCGAAGGACATCACCG 511
QY 697 CTGGCTGCACACCTATGAGGTGGTGGACTCTCTGTCAACACATCCCCAATCCAG 756
Db 512 CTGGCTGCACACCTATGAGGTGGTGGACTCTCTGTCAACACATCCCCAATCCAG 571
QY 757 CAAGATCTGCGGGTAAAGCTCAACCTGGCCCGCATGGCAGCAGAGCAAGAGAGAGCGT 816
Db 572 CAAGATCTGCGGGTAAAGCTCAACCTGGCCCGCATGGCAGCAGAGCAAGAGAGAGCGT 631
QY 817 CTTGTCAATCAGGCTGACCTGCAGAGGCGCAAGGCCCGAGCAGAGACCAAGCCACTGA 876
Db 632 CTTGTCAATCAGGCTGACCTGCAGAGGCGCAAGGCCCGAGCAGAGACCAAGCCACTGA 691
QY 877 GGAACCTGCGAGCTGGGAGAGAGAGCAGCGAGCCCGCTCAGGTCCAGGGTGAACGCG 936
Db 692 GGAACCTGCGAGCTGGGAGAGAGAGCAGCGAGCCCGCTCAGGTCCAGGGTGAACGCG 751
QY 937 CTTGAGACAGTCTGGCTGCTACACACATGCTGATGAGAACTGAGAGAGAGAAACA 996
Db 752 CTTGAGACAGTCTGGCTGCTACACACATGCTGATGAGAACTGAGAGAGAGAAACA 811
QY 997 CTAATTGATCTCGCCGGGATAGAAAATACAGTCCCAATTTGGGCTGCTCCCTTC 1056
Db 812 CTAATTGATCTCGCCGGGATAGAAAATACAGTCCCAATTTGGGCTGCTCCCTTC 871
QY 1057 CGTGGCCAGAAAGTCAGAACTGCGCGCGCGACCTCCCAATCCCACTGATTCGCTCCA 1116
Db 872 CGTGGCCAGAAAGTCAGAACTGCGCGCGCGACCTCCCAATCCCACTGATTCGCTCCA 931
QY 1117 TGAACCCGAAGCCATCCACATCCCAACCGAAAGCCCAAGGCTGAGACCCGGCTCTT 1176
Db 932 TGAACCCGAAGCCATCCACATCCCAACCGAAAGCCCAAGGCTGAGACCCGGCTCTT 991
QY 1177 CCACTTCTTGAACCCCAATCCCAAGTCTCAGAGCTCAGCAACGGCTCCGGGGCAC 1236
Db 992 CCACTTCTTGAACCCCAATCCCAAGTCTCAGAGCTCAGCAACGGCTCCGGGGCAC 1051

QY 1237 CAGAGCGGAGAGACATTTGTGAGTCCCAAGGCCAGAGGCAAGTGTGGGTGT 1296
Db 1052 CAGAGCGGAGAGACATTTGTGAGTCCCAAGGCCAGAGGCAAGTGTGGGTGT 1111
QY 1297 GGGCCAGTGGGCGAGAGGGGCAAGAAACAAGCCCTCTTGGGACCGGCATTCCTGCGGT 1356
Db 1112 GGGCCAGTGGGCGAGAGGGGCAAGAAACAAGCCCTCTTGGGACCGGCATTCCTGCGGT 1171
QY 1357 GTCCCCCTCCGCCACTGTGCTGACAGCCCGGCGCTCTCCCTCCCTAGCCCCCTCGG 1416
Db 1172 GTCCCCCTCCGCCACTGTGCTGACAGCCCGGCGCTCTCTCCCTAGCCCCCTCGG 1231
QY 1417 GCACAAAGAACAAAGACACGAGCCAAAGAGACACAGAGGCTGCGGAGCTTCAGAGC 1476
Db 1232 GCACAAAGAACAAAGACACGAGCCAAAGAGACACAGAGGCTGCGGAGCTTCAGAGC 1291
QY 1477 ACCACTGGCTCAGGTGGCCCTGTCTGTGGGGCGAGACCTGCGGGAGCTGCCGCTT 1536
Db 1292 ACCACTGGCTCAGGTGGCCCTGTCTGTGGGGCGAGACCTGCGGGAGCTGCCGCTT 1351
QY 1537 GGTGGTGTATGAGAGCCAGGCGGGGACAGCCGTCAGAGACATGAGACACACCA 1596
Db 1352 GGTGGTGTATGAGAGCCAGGCGGGGACAGCCGTCAGAGACATGAGACACCA 1411
QY 1597 TGAACATCACCACTTACCAACCACTTACCAAGCA 1633
Db 1412 TGAACATCACCACTTACCAACCACTTACCAAGCA 1448

RESULT 4
US-09-764-891-5902
; Sequence 5902, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5902
; LENGTH: 2379
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-5902

Query Match 42.1%; Score 783; DB 10; Length 2379;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1046 GGCCTCCCTTCCGTGGCCGAAAGTCAGAACTGCCCCCGGCACTTCAATCCACTCGA 1105
Db 840 GGCCTCCCTTCCGTGGCCGAAAGTCAGAACTGCCCCCGGCACTTCAATCCACTCGA 899
QY 1106 TCTGCTCCCATGAGCCGGAAGCCATCCAGATCCCAACGGAAGCCCAAGGGGTGGAC 1165
Db 900 TCTGCTCCCATGAGCCGGAAGCCATCCAGATCCCAACGGAAGCCCAAGGGGTGGAC 959
QY 1166 CCGGCTCTTTCACATTCCTTGAACCCCAATGCGCAAGGTCTCAGAGCTCCAGCAACGG 1225
Db 960 CCGGCTCTTTCACATTCCTTGAACCCCAATGCGCAAGGTCTCAGAGCTCCAGCAACGG 1019
QY 1226 CTCGGGGGACCCAGAGCGGAGCAAGCACTTTGTGAGTCCCAAGGCCCAAGGCGAAG 1285
Db 1020 CTCGGGGGACCCAGAGCGGAGCAAGCACTTTGTGAGTCCCAAGGCCCAAGGCGAAG 1079
QY 1286 AGTGTGGTGTGGGCGACCTGGCCAGAGGGGCAAGAAACAAGCCCCCTTGGGACCGGCC 1345
Db 1080 AGTGTGGTGTGGGCGACCTGGCCAGAGGGGCAAGAAACAAGCCCCCTTGGGACCGGCC 1139
QY 1346 ATCCCTCGGTGTCCCTCCGCCCACTGGCTGCAGGCCGGGCTCTTCCCTCCTA 1405


```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5
; OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8
; OTHER INFORMATION: SWISSPROT HIT: Q23977, EVALUE 7.70e-02
; OTHER INFORMATION: NT HIT: g14916432, EVALUE 9.00e-71
; OTHER INFORMATION: EST_HUMAN HIT: BF920913.1, EVALUE 3.00e-68
US-10-029-386-7990
```

```

Query Match
Best Local Similarity 100.0%; Pred. No. 2.2e-55; Length 598;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 280 AGGTGACAGCTTGGCCCTGAGCGCTGCTGGGCTCGGAAGGCGATCGAGAGTGATCGG 339
DB 228 AGGTGACAGCTTGGCCCTGAGCGCTGCTGGGCTCGGAAGGCGATCGAGAGTGATCGG 287
QY 340 GAGACAGCGCTGCGCGGCGGTGTCTGGGACCCCGACAGCTGCGTTGGCGGGCACCAT 399
DB 288 GAGACAGCGCTGCGCGGCGGTGTCTGGGACCCCGACAGCTGCGTTGGCGGGCACCAT 347
QY 400 AGGCGGAGACACCCCGG 416
DB 348 AGGCGGAGACACCCCGG 364
```

```

RESULT 12
US-10-016-157A-126
; Sequence 126, Application US/10016157A
; Publication No. US2002019220A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Heirve
; APPLICANT: Ghosh, Malayika
; APPLICANT: Liu, Chenshua
; TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and Pro
; FILE REFERENCE: DEX-0253
; CURRENT APPLICATION NUMBER: US/10/016,157A
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,717
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 250
; SOFTWARE: Patentln version 3.1
; SEQ ID NO 126
; LENGTH: 1950
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-016-157A-126
```

```

Query Match
Best Local Similarity 100.0%; Pred. No. 1.9e-55; Length 1950;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 281 GGTGACAGCTTGGCCCTGAGCGCTGCTGGGCTCGGAAGGCGATCGAGAGTGATCGGG 340
DB 838 GGTGACAGCTTGGCCCTGAGCGCTGCTGGGCTCGGAAGGCGATCGAGAGTGATCGGG 897
QY 341 AGACAGCGCTGCGCGGCGGTGTCTCGGACCCCGACAGCTGCGTTGGCGGGCACCAT 400
DB 898 AGACAGCGCTGCGCGGCGGTGTCTCGGACCCCGACAGCTGCGTTGGCGGGCACCAT 957
QY 401 GGCAGAGCACCCCGGA 417
DB 958 GGCAGAGCACCCCGGA 974
```

```

RESULT 13
US-09-993-966-14
; Sequence 14, Application US/09993966
; Publication No. US20030186232A1
; GENERAL INFORMATION:
; APPLICANT: ROMAN, MICHAEL
; TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NKD PROTEIN,
; FILE REFERENCE: 014024/0280733
; CURRENT APPLICATION NUMBER: US/09/993,966
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 60/252,884
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/291,109
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/325,571
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 14
; LENGTH: 134
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-993-966-14
```

```

Query Match
Best Local Similarity 100.0%; Pred. No. 7.2e-54; Length 134;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 282 GTGACAGCTTGGCCCTGAGCGCTGCTGGGCTCGGAAGGCGATCGAGAGTGATCGGA 341
DB 1 GTGACAGCTTGGCCCTGAGCGCTGCTGGGCTCGGAAGGCGATCGAGAGTGATCGGA 60
QY 342 GACAGCGCTGCGCGGCGGTGTCTGGGACCCCGACAGCTGCGTTGGCGGGCACCAT 401
DB 61 GACAGCGCTGCGCGGCGGTGTCTGGGACCCCGACAGCTGCGTTGGCGGGCACCAT 120
QY 402 GCCGAGACACCCCG 415
DB 121 GCCGAGACACCCCG 134
```

```

RESULT 14
US-10-029-386-21690
; Sequence 21690, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GI
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21690
; LENGTH: 133
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```

; FEATURE:
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5
; OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN HUMAN HIT: BF920913.1, EVALUE 4.00e-67
; OTHER INFORMATION: SWISSPROT HIT: P50570, EVALUE 2.60e+00
; OTHER INFORMATION: NT HIT: g14916432, EVALUE 4.00e-69
```


THIS PAGE IS BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: December 29, 2004, 16:07:39 ; Search time 12706 Seconds
(without alignments)
6918.901 Million cell updates/sec

Title: US-09-993-966-5

Perfect score: 1859
Sequence: 1 gaattcgcctctcaatcacg.....aaaaaaaaaagcggcgccg 1859

Scoring table: OLIGOBLAST
Gapop 60.0 , Gapext 60.0

Searched: 4526729 seqs, 23644849745 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_ste:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1786	96.1	1786	AY061883	AY061883 Homo sapi
2	1763	94.8	2604	BC051288	BC051288 Homo sapi
3	1610	86.6	1788	AB062886	AB062886 Homo sapi
4	1438	77.4	1438	AF358135	AF358135 Homo sapi
5	1336	71.9	1438	AR405917	AR405917 Sequence
6	812	43.7	1053	CQ721681	CQ721681 Sequence
7	783	42.1	163319	HSAB303140	HAB303140 Homo sapi
8	783	42.1	169773	AC007608	AC007608 Homo sapi
9	783	42.1	170791	AC007334	AC007334 Homo sapi
10	528	28.4	178790	AC145238	AC145238 Pan trogl
11	528	28.4	207486	AC145175	AC145175 Pan trogl
12	137	7.4	1950	AX557538	AX557538 Sequence
13	136	7.3	2340	CQ741468	CQ741468 Sequence
14	124	6.7	194869	AC145003	AC145003 Papio anu
15	124	6.7	198189	AC145043	AC145043 Papio anu
16	17	3.8	181401	AC146663	AC146663 Sus scrofa
17	68	3.7	2469	C0843478	C0843478 Sequence
18	68	3.7	2469	AK124375	AK124375 Homo sapi
19	65	3.5	199277	AC145445	AC145445 Canis fam

20	58	3.1	670	6	AX530283	AX530283 Sequence
21	57	3.1	417	6	AX530292	AX530292 Sequence
22	57	3.1	510	6	AX530281	AX530281 Sequence
23	56	3.0	636	6	AX530289	AX530289 Sequence
24	56	3.0	651	6	AX530272	AX530272 Sequence
25	56	3.0	686	6	AX530282	AX530282 Sequence
26	56	3.0	2048	4	AJ715983	AJ715983 Equus cab
27	56	3.0	3951	10	AF304853	AF304853 Rattus no
28	56	3.0	208061	2	AC145254	AC145254 Bos tauru
29	55	3.0	547	6	AX530274	AX530274 Sequence
30	55	3.0	1245	6	AX530269	AX530269 Sequence
31	54	2.9	1874	6	AX528996	AX528996 Sequence
32	54	2.9	1874	6	AX529000	AX529000 Sequence
33	54	2.9	2437	3	AJ620505	AJ620505 Lacanobia
34	54	2.9	2461	8	MS1428052	AJ428052 Malassezi
35	54	2.9	2605	6	AX529022	AX529022 Sequence
36	53	2.9	699	3	SFR488181	AJ488181 Spodopter
37	53	2.9	711	8	VAL457214	AJ457214 Vileum a1
38	53	2.9	1434	6	CQ792377	CQ792377 Sequence
39	53	2.9	1434	6	AX456458	AX456458 Sequence
40	53	2.9	3003	10	AB090950	AB090950 Cania por
41	53	2.9	3281	6	AX708831	AX708831 Sequence
42	52	2.8	577	6	AX530278	AX530278 Sequence
43	51	2.7	2241	5	SSA55695	AJ55695 Salmo sal
44	50	2.7	226	6	AX530284	AX530284 Sequence
45	50	2.7	1401	6	AX172351	AX172351 Sequence

ALIGNMENTS

RESULT 1	AY061883	1786 bp	mRNA	linear	PRI 26-DEC-2001
LOCUS	AY061883				
DEFINITION	Homo sapiens naked protein (NKD)				
ACCESSION	AY061883				
VERSION	AY061883.1	GI:17978536			
KEYWORDS					
SOURCE					
ORGANISM	Homo sapiens (human)				
REFERENCE					
AUTHORS	1 (bases 1 to 1786)				
TITLE	Yan, D., Wiesmann, M., Rohan, M., Chan, V., Jefferson, A. B., Guo, L., Sakamoto, D., Caothien, R. H., Fuller, J. H., Reinhard, C., Garcia, P. D., Randazzo, F. M., Escobedo, J., Pantl, W. J. and Williams, L. T.				
JOURNAL	Elevated expression of axin2 and nkhd mRNA provides evidence that Wnt/beta-catenin signaling is activated in human colon tumors				
PUBMED	Proc. Natl. Acad. Sci. U.S.A. 98 (26), 14973-14978 (2001)				
REFERENCE	11752446				
AUTHORS	2 (bases 1 to 1786)				
TITLE	Rohan, M., Chan, V., Yan, D., Pantl, W. J. and Williams, L. T.				
JOURNAL	Direct Submission				
FEATURES	Submitted (05-NOV-2001) Cancer Biology, Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608, USA				
source	Location/Qualifiers				
gene	1..1786				
CDS	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	1..1786				
	/gene="NKD"				
	159..1571				
	/gene="NKD"				
	/note="HNKD"				
	/codon_start=1				
	/product="naked protein"				
	/protein_id="AAJ32374.1"				
	/db_xref="GI:17978537"				
	/translation="MGKLSKPAVCRRRESPGDSFPAVSAAMARKGIEEMIGRQPC				
	GGVSGPRLAIGTIGSTREIVAGDVLRDTLSEEDDFLEVALPEKTDGSGDE				
	KMERVSEPCGSKQKPEELQCDVSMEDSDQEWTFLLDNNNGKVRREDTLSL				
	HTIYEVDSVSNHSPTSSKMLRVKLTVAAPDGSGSKSVLVNQADLQSGARRARAEKPPPE				

ORIGIN

65 AGTCGGGCGCCAGCAGCGCTACGCATCGGCCATTACGCAACA 125

126	CTCGGGGGCTCTTCGGAAGAGAGCAAAAGGAGCGCCAGAGCCCGCGGCGCGGGCG	185
61	CTCGGGGGCTCTTCGGAAGAGAGCAAAAGGAGCGCCAGAGCCCGCGGCGCGGGCG	120
186	CATGGCTTTAGGAACGCTCCCGGCGCGCAGCCCCAGCATGGAGAACTTCACTCAAGC	245
121	CATGGCTTTAGGAACGCTCCCGGCGCGCAGCCCCAGCATGGAGAACTTCACTCAAGC	180
246	CGGCGCGCGTGTGCAAGCGCAGAGAGAGCCCGAAGGTGACAGCTTCGCCCTGAGCGCTG	305
181	CGGCGCGCGTGTGCAAGCGCAGAGAGAGCCCGAAGGTGACAGCTTCGCCCTGAGCGCTG	240
306	CCTGGGCTCGGAAGGCAATCGAAGAGTGAATCGGGAAGACAGCGCGCCCGGAGCGTGTCT	365
241	CTTGGGCTCGGAAGGCAATCGAAGAGTGAATCGGGAAGACAGCGCGCCCGGAGCGTGTCT	300
366	CGGGAACCCCGACAGCTGCGATTGCGCGGACCATAGCCGAAGACCCGGAGCTCGTG	425
301	CGGGAACCCCGACAGCTGCGATTGCGCGGACCATAGCCGAAGACCCGGAGCTCGTG	360
426	GCGACGTGTGAGAGACAGCTCAGCGGAGAAAGAGAGACGACTTTCGCTGGAAGTGG	485
361	GCGACGTGTGAGAGACAGCTCAGCGGAGAAAGAGAGACGACTTTCGCTGGAAGTGG	420
486	CCCGGCTCCCTGAGAAAGCTGACGGGCGCGGCAACGGAGATGAGAAAGAAATGAGAGAG	545
421	CCCGGCTCCCTGAGAAAGCTGACGGGCGCGGCAACGGAGATGAGAAAGAAATGAGAGAG	480
546	TGAGCGAACCCTGCGCCAGGCTCCAAGAAAGCAGTGAAGTTGAAGAGCTCCAGTGCAGC	605
481	TGAGCGAACCCTGCGCCAGGCTCCAAGAAAGCAGTGAAGTTGAAGAGCTCCAGTGCAGC	540
606	TGTTCATGGAAGAGACAGCCGCGAGAGTGAACCTTCACTCTGTATGACTTTGACAA	665
541	TGTTCATGGAAGAGACAGCCGCGAGAGTGAACCTTCACTCTGTATGACTTTGACAA	600
666	ACGGCAAGGTCACCCGAGAGGACATTCACAGCTTGCGACACACATCTATAGAGTGGTG	725
601	ACGGCAAGGTCACCCGAGAGGACATTCACAGCTTGCGACACACATCTATAGAGTGGTG	660
726	ACTCTCTGTCAACCACTCCCAACATTCAGCAGAGATCTCGGGTAAAGCTCACCGTGG	785
661	ACTCTCTGTCAACCACTCCCAACATTCAGCAGAGATCTCGGGTAAAGCTCACCGTGG	720
786	CCCCCGATGGCAGCCAGAGCAAGAGAGCGTCTTGTCAATCAGACTGACCTTGACAGCG	845
721	CCCCCGATGGCAGCCAGAGCAAGAGAGCGTCTTGTCAATCAGACTGACCTTGACAGCG	780
846	CAAGCCCCCGAGCAGAGACCAAGCCCACTAGAGACCTGCGGAGCTGGGAGAAAGACAGC	905
781	CAAGCCCCCGAGCAGAGACCAAGCCCACTAGAGACCTGCGGAGCTGGGAGAAAGACAGC	840
906	GAGGCCCGCTCAGGTTCCAGGGTGAACAGCGCGCTGAGAGACATCTGGCTGTACACCAATT	965
841	GAGGCCCGCTCAGGTTCCAGGGTGAACAGCGCGCTGAGAGACATCTGGCTGTACACCAATT	900
966	GCGTATGAGAAATCATGAGAGAGAAACCACTATAGATCTCGCGGAGTAAAGAACT	1025

[illegible]

RESULT 2	
LOCUS	BC051288
DEFINITION	BC051288 Homo sapiens naked cuticle homolog 1 (<i>Drosophila</i>), mRNA (cDNA clone)
ACCESSION	BC051288
VERSION	BC051288.1
KEYWORDS	MGC.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. (bases 1 to 2604)
AUTHORS	Strasberg,L.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,

Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, J.T., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Uedini, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mulhally, S.J., Bosak, S.A., McEwan, P.J., McKernan, R.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huiyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalski, U., Smailus, D.E., Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932
2 (bases 1 to 2604)
Strausberg, R.
Direct Submission
Submitted (25-APR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgaps-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mgc@nih.gov
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R., Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantipod, S., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
Series: IRML Plate: 47 Row: O Column: 23
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 20127660.

Location/Qualifiers
1. 2604
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:59666 IMAGE:6646089"
/tissue_type="Ovary, teratocarcinoma"
/clone_lib="NIH MGC_109"
/lab_host="DH10B-R"
/note="Vector: pOT87"
1. 2604
/gene="NKD1"
/db_xref="locusID:85407"
/db_xref="MTM:607851"
225. 1637
/gene="NKD1"
/codon_start=1
/product="naked cuticle homolog 1"
/protein_id="AAH51288.1"
/db_xref="GI:30410966"

ORIGIN

Query Match 94.8%; Score 1763; DB 9; Length 2604;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1763; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

/db_xref="MTM:607851"
/transl="WGLKHSKPAVCKRBSPEGDSFPAVSAMARKIEWIRORCP
GGVSGRORLNGTTRSTRELVDVRLTSEEDDDRLLEALPEKTDGSGDE
KMERSEPCPGSKKOLKEELOCDSMEDSROEWTFTLYPDNNGKTVREITSL
HTREYVDSVSNHSPTSSKMLRYKLVTAAPDGSOSKSVLVNADLOSARPRATPTE
DLRSMKKRRLPRFGDSRLBOSGVYHCHENIERRHVLDLAGIENTSOFSGS
PSVAOKSELPRPSTNTPRSRSHPEALINHPKRPQVDPAFSFPLTPAKVSELDOR
LRKTDGSKHPVSPFAOGKSVGVHVAAGKRRPGLPATIPAVSSAHILASPALIP
SLAPLHKHKKHRAKSKOQCGRLDAPLASGGVLRREHLRLPALVVESSOGAPVO
RHHHHHHHHHHHHFYQT"

66 AGTCGGGCGCGGCGAGCGGCGGAGAGCGCTCCGCGCGCGCTCGGCTCGG 125
67 AGTCGGGCGCGGCGAGCGGCGGAGAGCGCTCCGCGCGCGCTCGGCTCGG 126
126 CTCGGGGGCTGCTTTCGGAGAGAGAGCCGCAAGAGCGCCGCGCGCGCGG 185
127 CTCGGGGGCTGCTTTCGGAGAGAGAGCCGCAAGAGCGCGCCGCGCGCGG 186
186 CATGCTTAGAGAGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 245
187 CATGCTTAGAGAGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 246
246 CGGCGCGCGCTGTCAGAGCGGCGGAGAGCGCGGAGAGTGAACGCTTCGCGTGC 305
247 CGGCGCGCGCTGTCAGAGCGGCGGAGAGCGCGGAGAGTGAACGCTTCGCGTGC 306
306 CCGGCGCGCGGAGAGCGGCGGAGAGTGAACGCGCTGCCTCGCGCGCTGCT 365
307 CCGGCGCGCGGAGAGCGGCGGAGAGTGAACGCGCTGCCTCGCGCGCTGCT 366
366 CGGAGACCGCGAGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 425
367 CGGAGACCGCGAGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 426
426 GCGAGCGTGAAGAGCAGCGTCAAGCGAGAGAGAGAGCGCTTCGCGTGAAGTGG 485
427 GCGAGCGTGAAGAGCAGCGTCAAGCGAGAGAGAGAGCGCTTCGCGTGAAGTGG 486
486 CCCTGCTCTCTGAAGAGCTGAGCGGCTGGCGGCGGAGTGAAGAGAGAGAG 545
487 CCCTGCTCTCTGAAGAGCTGAGCGGCTGGCGGCGGAGTGAAGAGAGAGAG 546
546 TGAAGCAACCTGCGCGAGAGTCAAGAGCGTGAAGTGAAGAGCTCAAGTGCAGC 605
547 TGAAGCAACCTGCGCGAGAGTCAAGAGCGTGAAGTGAAGAGCTCAAGTGCAGC 606
606 TGTCATGAGAGAGAGCGGCGGAGAGTGAAGTGAAGTGAAGAGCTTCAGCTTGAACACA 665
607 TGTCATGAGAGAGAGCGGCGGAGAGTGAAGTGAAGTGAAGAGCTTCAGCTTGAACACA 666
666 ACGGCAAGTCAACCGAGAGAGATCAAGAGCGTTCGAGCAACATTAAGAGTGTGG 725
667 ACGGCAAGTCAACCGAGAGAGATCAAGAGCGTTCGAGCAACATTAAGAGTGTGG 726
726 ACTCTCTGTGAACCACTCCCAACATCAAGCAAGTGTGGAGTGAAGAGTGTGG 785
727 ACTCTCTGTGAACCACTCCCAACATCAAGCAAGTGTGGAGTGAAGAGTGTGG 786
786 CCGCGAGTGAAGAGCGGAGAGAGAGAGAGTGTGTCAATCAAGAGTGTGTGAAGAGC 845
787 CCGCGAGTGAAGAGCGGAGAGAGAGAGAGTGTGTCAATCAAGAGTGTGTGAAGAGC 846
846 CAAAGCCCGAGC 905
847 CAAAGCCCGAGC 906

```
QY 906 GAGCCCCGCTCAGGTTCCAGGGTGAAGACCCGCTGAGCAGTGTGCTGACCAACATT 965
Db 907 GAGCCCCGCTCAGGTTCCAGGGTGAAGACCCGCTGAGCAGTGTGCTGACCAACATT 966
QY 966 GCGTAGATGAGAACATCGAGAGAGAAACACTATTAGTCTCGCGGGATAGAAAAT 1025
Db 967 GCGTAGATGAGAACATCGAGAGAGAAACACTATTAGTCTCGCGGGATAGAAAAT 1026
QY 1026 ACAAGTCCCAATTTGGGCTGGCTCCCTCCCTGGGCGCCAGAAATCGAAATGCCCCC 1085
Db 1027 ACAAGTCCCAATTTGGGCTGGCTCCCTCCCTGGGCGCCAGAAATCGAAATGCCCCC 1086
QY 1086 GCACCTCCATCCCATCTGATCTGCTCCCATGAGCGGAGGCATCCATCCCAACACC 1145
Db 1087 GCACCTCCATCCCATCTGATCTGCTCCCATGAGCGGAGGCATCCATCCCAACACC 1146
QY 1146 GAAAGCCCCAAGGCGTGGACCCGGCTCTTCCATTCCATTGACACCCCAATGCGCAAG 1205
Db 1147 GAAAGCCCCAAGGCGTGGACCCGGCTCTTCCATTCCATTGACACCCCAATGCGCAAG 1206
QY 1206 TCTTGAAGCTCCAGCAACGGCTCCGGGGCACCCAGAGCGGAGGCACTTTGTAGGT 1265
Db 1207 TCTTGAAGCTCCAGCAACGGCTCCGGGGCACCCAGAGCGGAGGCACTTTGTAGGT 1266
QY 1266 CCCCCAAGGCCCAAGGCAAGTGTGGGTGTGGGCACTGTGGCCAGAGGGGCAAGAAACA 1325
Db 1267 CCCCCAAGGCCCAAGGCAAGTGTGGGTGTGGGCACTGTGGCCAGAGGGGCAAGAAACA 1326
QY 1326 AGCCCCCTCTGGGACCCGCACTCCCTGCGGTGTCCCTCCGACCTGGCTGCGCAGCC 1385
Db 1327 AGCCCCCTCTGGGACCCGCACTCCCTGCGGTGTCCCTCCGACCTGGCTGCGCAGCC 1386
QY 1386 CCGGCCCCCTCCCTCCCTAGCCCCCTCGGGGCAAGAAAGCAACACACCGAGCCAAAG 1445
Db 1387 CCGGCCCCCTCCCTCCCTAGCCCCCTCGGGGCAAGAAAGCAACACACCGAGCCAAAG 1446
QY 1446 AGAGCCAGCAGGCGTGCCTGGGGCTCGAGGCAACCACTGAGTGGCCCTGTCTTG 1505
Db 1447 AGAGCCAGCAGGCGTGCCTGGGGCTCGAGGCAACCACTGAGTGGCCCTGTCTTG 1506
QY 1506 GCGGGAGGACCTGCGGGAGCTGCGGCTTGTGTATGAGAGCCAGGCGCGGAGC 1565
Db 1507 GCGGGAGGACCTGCGGGAGCTGCGGCTTGTGTATGAGAGCCAGGCGCGGAGC 1566
QY 1566 CCGTCCAGAGACATGAGCACCAACCAACATGAACTCAACCACTTACCAACACTTCT 1625
Db 1567 CCGTCCAGAGACATGAGCACCAACCAACATGAACTCAACCACTTACCAACACTTCT 1626
QY 1626 ACCAGACATAGAGCCCTCCCGAGGGCCCACTGTCATATGAGAGACCCCAACCCCGA 1685
Db 1627 ACCAGACATAGAGCCCTCCCGAGGGCCCACTGTCATATGAGAGACCCCAACCCCGA 1686
QY 1686 CACCACAGGCAATTAATTCATTAATAATTTGTTATTAATGATTAATTTGTTAATA 1745
Db 1687 CACCACAGGCAATTAATTCATTAATAATTTGTTATTAATGATTAATTTGTTAATA 1746
QY 1746 AATTATTTACTCCATTAATTTAGTACCTTACATGAGAAATCTTAGAGAAACAG 1805
Db 1747 AATTATTTACTCCATTAATTTAGTACCTTACATGAGAAATCTTAGAGAAACAG 1806
QY 1806 AACTAACTTTATTTATATGTT 1828
Db 1807 AACTAACTTTATTTATATGTT 1829
```

```
RESULT 3
LOCUS AB062886 1788 bp mRNA linear PRI 20-OCT-2001
DEFINITION Homo sapiens mRNA for Dvl-binding protein NKD1, complete cds.
ACCESSION AB062886
VERSION AB062886.1 GI:16303259
KEYWORDS
SOURCE Homo sapiens (human)
```

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Katoh,M.
TITLE Molecular cloning, gene structure, and expression analyses of NKD1
and NKD2
JOURNAL Int. J. Oncol. 19 (5), 963-969 (2001)
MEDLINE 21490203
PUBMED 11604995
REFERENCE 2 (bases 1 to 1788)
AUTHORS Katoh,M.
TITLE Direct Submission
JOURNAL Submitted (08-JUN-2001) Masaru Katoh, National Cancer Center
Research Institute, Genetics and Cell Biology Section: Tsukiji
5-chome, Chuo-ku, Tokyo 104-0045, Japan (E-mail:mkatoh@ncc.go.jp,
Tel:81-3-3542-2511)
FEATURES
source
1..1788
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue="human fetal tissues"
1..1788
/gene="NKD1"
5..1418
/gene="NKD1"
/codon_start=1
/product="Dvl-binding protein NKD1"
/protein_id="BAB70500.1"
/db_xref="GI:16303260"
/translation="MGKLSKPAVACKRESPPGDSFAVSAAMARKGIEBMLGRCPC
GCVSGRQLRLAGTIGSRRELIVDLRLTLSEEDDRLRLVLPKDTLDGSDR
KKMERVSEPCPSKQKLEBELQDYSMEDSDQETPTLYLDNNKXLRBETSL
HTLYEVLDSSVNSPTSSFMRLRYKLVADPSQSKSVLVNQLDLSAPRAETKTE
DLRSWEKQRAPLRFQDSRLBQSGYVBNICVENIRRNHYLDLAGIENYTSQFGGS
PVSQKSELPRPTNSRSHPEALIHPIHRKPDQVDPSFHLDTPLAKVSELQOR
LRGTODGSHFVRS PKAOGKSVGVNARGARKPLGPALIPAVSPSAILAASPLLP
SLAPLGKHKHKAKEKSGOQCRGLQPLASGCVLRLRHLRELPLLVVESAGQPVQ
RHHNNHHHHHHHHFYQT"
ORIGIN
Query Match 86.6%; Score 1610; DB 9; Length 1788;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1610; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 219 CCAGCATGGGAAATCTTCACTCCAAAGCGGCGCGTGTGCAAGCGGAGGAGCCCGG 278
Db 1 CCAGCATGGGAAATCTTCACTCCAAAGCGGCGCGCGTGTGCAAGCGGAGGAGCCCGG 60
QY 279 AAGGTACAGCTTCGCGGTGAGGCGTGCCTGGGCTGGAAAGGCATCGAGAGTGATCG 338
Db 61 AAGGTACAGCTTCGCGGTGAGGCGTGCCTGGGCTGGAAAGGCATCGAGAGTGATCG 120
QY 339 GGAAGACGCGCTCCCGGCGGTGTCTCGGGAACCCGACAGACTGCGGTTGGCGGGACCA 398
Db 121 GGAAGACGCGCTCCCGGCGGTGTCTCGGGAACCCGACAGACTGCGGTTGGCGGGACCA 180
QY 121 GGAAGACGCGCTCCCGGCGGTGTCTCGGGAACCCGACAGACTGCGGTTGGCGGGACCA 180
QY 399 TAGGCGGAAGCACCCGGAGCTCTGTGGGCGAGCTGTGAGAGACAGCTTACGAGAGAG 458
Db 181 TAGGCGGAAGCACCCGGAGCTCTGTGGGCGAGCTGTGAGAGACAGCTTACGAGAGAG 240
QY 459 AGAGAGACGACTTTCGCTGGAAGTGGCCCTGCTCTGAGAGACTGAGCGGCTGGGCA 518
Db 241 AGAGAGACGACTTTCGCTGGAAGTGGCCCTGCTCTGAGAGACTGAGCGGCTGGGCA 300
QY 519 GCGGATGAGAGAGATGAGAGATGAGAGCAACCTTGCAGGCTCCAGAAAGCAGC 578
Db 301 GCGGATGAGAGAGATGAGAGATGAGAGCAACCTTGCAGGCTCCAGAAAGCAGC 360
QY 579 TGAAGTTTGAAGAGCTTCAGTGCAGCGTGTCCATGAGAGAGACAGCGGCGAGAGTGGA 638
Db 361 TGAAGTTTGAAGAGCTTCAGTGCAGCGTGTCCATGAGAGAGACAGCGGCGAGAGTGGA 420
```

QY 639 CCTTCACTCTGTATGACTTTTGACACACGCGCAAGTCCCGGAGAGACATCAACAGCT 698
 DB 421 CCTTCACTCTGTATGACTTTTGACACACGCGCAAGTCCCGGAGAGACATCAACAGCT 480
 QY 699 TGTGACACACATTTATGAGGTGTGTGAGCTCTGTCAACACATCCCGCAATCCAGCA 758
 DB 481 TGTGACACACATTTATGAGGTGTGTGAGCTCTGTCAACACATCCCGCAATCCAGCA 540
 QY 759 AGATGCTGCGGGTAAAGCTCACGTGACCCCGATGAGCAGCAGACAGAGAGAGCGTCC 818
 DB 541 AGATGCTGCGGGTAAAGCTCACGTGACCCCGATGAGCAGCAGACAGAGAGAGCGTCC 600
 QY 819 TTGTCAATCAGGCTGACTGTCAGAGCGGACCGGACGAGAGACCAAGCCACTAGAG 878
 DB 601 TTGTCAATCAGGCTGACTGTCAGAGCGGACCGGACGAGAGACCAAGCCACTAGAG 660
 QY 879 ACCTGCGGAGCTGGAGAGAGAGAGAGCGGCTCAGGTTCCAGGTTGACAGCGGCC 938
 DB 661 ACCTGCGGAGCTGGAGAGAGAGAGAGCGGCTCAGGTTCCAGGTTGACAGCGGCC 720
 QY 939 TGGAGCAGTGTGCTGTACACCATTTGCTGATGAGAACATCAGAGAGAGAACACT 998
 DB 721 TGGAGCAGTGTGCTGTACACCATTTGCTGATGAGAACATCAGAGAGAGAACACT 780
 QY 999 ACTTAATCTCTGCGGGATGAGAAACTACACGTCCCAATTTGGGCTTGGCTCCCTTCG 1058
 DB 781 ACTTAATCTCTGCGGGATGAGAAACTACACGTCCCAATTTGGGCTTGGCTCCCTTCG 840
 QY 1059 TGGCCCGAAGTGAACCTGCCCCCGGACCTCAATCCACCTGATCTGCTCCCATG 1118
 DB 841 TGGCCCGAAGTGAACCTGCCCCCGGACCTCAATCCACCTGATCTGCTCCCATG 900
 QY 1119 AGCGGAGAGCATTCACATCCCAACCGAAGCCCAAGGCTGAGACCGGCTCTTCC 1178
 DB 901 AGCGGAGAGCATTCACATCCCAACCGAAGCCCAAGGCTGAGACCGGCTCTTCC 960
 QY 1179 ACTTCTTGAACCCCAATGCGCAAGGCTCTGAGCTTCAGAACAGGCTCGGAGCAACC 1238
 DB 961 ACTTCTTGAACCCCAATGCGCAAGGCTCTGAGCTTCAGAACAGGCTCGGAGCAACC 1020
 QY 1239 AGAGCGGAGCAACGCACTTTGTAGGTCCCCAAGGCGGCAAGAGTGTGGTGG 1298
 DB 1021 AGAGCGGAGCAACGCACTTTGTAGGTCCCCAAGGCGGCAAGAGTGTGGTGG 1080
 QY 1299 GCGACGTGCGGAGGCGGCAAGAACAGGCTCTGAGACCGGCTCCGCTGCTG 1358
 DB 1081 GCGACGTGCGGAGGCGGCAAGAACAGGCTCTGAGACCGGCTCCGCTGCTGCTG 1140
 QY 1359 CCCCCTCGGCACTGTGCTGCGACGCGGCTCTCTCCCTCCCTAGCCCTCGGCG 1418
 DB 1141 CCCCCTCGGCACTGTGCTGCGACGCGGCTCTCTCCCTCCCTAGCCCTCGGCG 1200
 QY 1419 ACAAGAGCAACGCAACCGGAGCAAGAGAGCGGCTGCGGAGCTTCAGAGCAC 1478
 DB 1201 ACAAGAGCAACGCAACCGGAGCAAGAGAGCGGCTGCGGAGCTTCAGAGCAC 1260
 QY 1479 CATGCGCTGAGTGTGCTGTCTGAGGCGGAGCACTGCGGAGCTGCGGCTTGG 1538
 DB 1261 CATGCGCTGAGTGTGCTGTCTGAGGCGGAGCACTGCGGAGCTGCGGCTTGG 1320
 QY 1539 TGGTGTATGAGAGCGGAGCGGAGCGGCTCAGAGACATGAGACACACCAACCATG 1598
 DB 1321 TGGTGTATGAGAGCGGAGCGGAGCGGCTCAGAGACATGAGACACACCAACCATG 1380
 QY 1599 AAGATCAACCACTTACCACTTTTACAGAGACATGAGAGCTTCCCGAGGCGCCAC 1658
 DB 1381 AAGATCAACCACTTACCACTTTTACAGAGACATGAGAGCTTCCCGAGGCGCCAC 1440
 QY 1659 CTGCGATATGAGAGCGGAGCGGAGCGGACCAACCAAGGACATTAATTATTAATTG 1718
 DB 1441 CTGCGATATGAGAGCGGAGCGGAGCGGACCAACCAAGGACATTAATTATTAATTG 1500

QY 1719 TTATATGATGATATTTATTTATTAATTAATTGTACTCACTAATATTTAGTACCT 1778
 DB 1501 TTATATGATGATATTTATTTATTAATTAATTGTACTCACTAATATTTAGTACCT 1560
 QY 1779 ACATGTAGAATCTATGAGAAACAGACAACTAACTTTATTTATTTATTT 1828
 DB 1561 ACATGTAGAATCTATGAGAAACAGACAACTAACTTTATTTATTTATTT 1610
 RESULT 4
 AF358135 1438 bp mRNA linear PRI 29-MAY-2001
 LOCUS Homo sapiens naked cuticle-1 (NKD1) mRNA, complete cds.
 DEFINITION
 ACCESSION AF358135
 VERSION AF358135.1 GI:14211713
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1438)
 AUTHORS Wharton,K.A. Jr., Zimmermann,G., Rousset,R. and Scott,M.P.
 TITLE Vertebrate proteins related to Drosophila Naked Cuticle bind Dishevelled and antagonize Wnt signaling
 JOURNAL Dev. Biol. 234 (1), 93-106 (2001)
 MEDLINE 2125530
 PUBMED 11356022
 REFERENCE 2 (bases 1 to 1438)
 AUTHORS Wharton,K.A. Jr., Zimmermann,G. and Scott,M.P.
 TITLE Direct Submission
 JOURNAL Submitted (08-MAR-2001) Pathology and Molecular Biology, University of Texas Southwestern Medical School, 5323 Harry Hines Blvd., Dallas, TX 75390-9072, USA
 FEATURES
 source location/Qualifiers
 1..1438
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="16"
 /map="16q12"
 1..1438
 /gene="NKD1"
 8..1420
 /note="NKD1"
 /note="contains EFX domain; similar to Drosophila melanogaster naked cuticle (nkd) and to Mus musculus Nkd1"
 melanogaster=1
 /product="naked cuticle-1"
 /protein_id="AAK57484.1"
 /db_xref="GI:14211714"
 /translation="MGKLSKRAVAVCKRESPEGDSFVSAAMARKGIEBIGRORCPGGVSGPROLYAGTIGSTRELVDVLRDTLSSEEDPFLVLAIPREKTDGSGDPEKKMERVSEPCGSKKOLREELQDSMEDRSQEMTFTLYDNNNGKVTREDITSLHTIYEVLDSSVNHSPTSKKLYKLVLRDVSQSKSVLVNODLQASRPATKPTEDLSWEEKDQALRFQDSRLQSGCTYNNVDNIRRNKYLIDLAGEITSGPQESPSVAQKSELPRTSNPTRSNSHPEAIDIPHRQVDVDFPFLDPLAKVLELQRLRGTDGSKIFVRSYKQSKSVGVGVNAGARANKRPVGPALPAVSSAMLAAPALPRLSLRLHKKHKNHAKESQDQSCRGDQRLASGQVLVREHDLREPLALVVVSSQAGPQVRHNNHNNHNNHNNHNYDT"
 ORIGIN
 Query Match 77.4%; Score 1438; DB 9; Length 1438;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 217 CCCCAGATGGGAAACTTACTCCAAAGCGGCGCGCTGTGCAAGGCGAGAGAGCCC 276
 DB 1 CCCCAGATGGGAAACTTACTCCAAAGCGGCGCGCTGTGCAAGGCGAGAGAGCCC 60
 QY 277 GGAAGTGAAGCTTGGCCGTGAGCGCTGAGGCTCGGAAGGCAATCGAGAGTGGAT 336
 DB 61 GGAAGTGAAGCTTGGCCGTGAGCGCTGAGGCTCGGAAGGCAATCGAGAGTGGAT 120


```

Db      481 |CTTCTGACACCATCTATGAGGTGGTGGATCTCTCTGTCAACCACTTCCCAATCCAG 540
Qy      757 |CAAGATCTCGGGTAAAGCTCACCGTGGCCCCGATGGACGACGAGCAAGAGAGCGT 816
Db      541 |CAAGATCTCGGGTAAAGCTCACCGTGGCCCCGATGGACGACGAGCAAGAGAGCGT 600
Qy      817 |CTTGTCAATCAAGCTGACCTGCAAGGCGCAAGGCCCCGAGCAAGACCAAGCCACTGA 876
Db      601 |CTTGTCAATCAAGCTGACCTGCAAGGCGCAAGGCCCCGAGCAAGACCAAGCCACTGA 660
Qy      877 |GACCTGCGAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 936
Db      661 |GACCTGCGAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Qy      937 |CTTGAAGCACTGGCTGCTACCACTTGGATGAGAACTGAGAGAGAGAGAGAGAGAG 996
Db      721 |CTTGAAGCACTGGCTGCTACCACTTGGATGAGAACTGAGAGAGAGAGAGAGAGAG 780
Qy      997 |CTTGAAGCACTGGCTGCTACCACTTGGATGAGAACTGAGAGAGAGAGAGAGAGAG 1056
Db      781 |CTTGAAGCACTGGCTGCTACCACTTGGATGAGAACTGAGAGAGAGAGAGAGAGAG 840
Qy      1057 |CTTGAAGCACTGGCTGCTACCACTTGGATGAGAACTGAGAGAGAGAGAGAGAGAG 1116
Db      841 |CTTGAAGCACTGGCTGCTACCACTTGGATGAGAACTGAGAGAGAGAGAGAGAGAG 900
Qy      1117 |TGAAGCCGGAAGCCATCCACATCCACACGAGAGAGAGAGAGAGAGAGAGAGAGAG 1176
Db      901 |TGAAGCCGGAAGCCATCCACATCCACACGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
Qy      1177 |CACTTCTTGAACAGCCCAATCCGCAAGGTCTCAAGGTCTCAAGCAAGGTCTCGAG 1236
Db      961 |CACTTCTTGAACAGCCCAATCCGCAAGGTCTCAAGGTCTCAAGCAAGGTCTCGAG 1020
Qy      1237 |CAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1296
Db      1021 |TCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Qy      1297 |GAGGCAAGGTGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1356
Db      1081 |GAGGCAAGGTGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Qy      1357 |GTCCCTCTCGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1416
Db      1141 |GTCCCTCTCGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Qy      1417 |GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476
Db      1201 |GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Qy      1477 |ACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1536
Db      1261 |ACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Qy      1537 |GGTGGTGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1596
Db      1321 |GGTGGTGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Qy      1597 |TGAACATCACACCATTAACCACTTCTACCAAGACATAGAGAGAGAGAGAGAGAG 1654
Db      1381 |TGAACATCACACCATTAACCACTTCTACCAAGACATAGAGAGAGAGAGAGAGAG 1438

```

```

RESULT 6
CQ721681      1053 bp   DNA      linear   PAT 03-FEB-2004
LOCUS         CQ721681
DEFINITION    Sequence 7615 from Patent WO02068579.
ACCESSION     CQ721681
VERSION       CQ721681.1 GI:42282538
KEYWORDS
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens

```

```

REFERENCE
AUTHORS      Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
TITLE        Kites, such as nucleic acid arrays, comprising a majority of
              humenexons or transcripts, for detecting expression and other uses
              thereof
JOURNAL      Patent: WO 02068579-A 7615 06-SEP-2002;
              PE Corporation
FEATURES
source       1..1053
              /organism="Homo sapiens"
              /mol_type="unassigned DNA"
              /db_xref="taxon:9606"
ORIGIN
Query Match      43.7%; Score 812; DB 6; Length 1053;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 812; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1
Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
Kites, such as nucleic acid arrays, comprising a majority of
humenexons or transcripts, for detecting expression and other uses
thereof
Patent: WO 02068579-A 7615 06-SEP-2002;
PE Corporation
1..1053
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
Qy      481 |AGTGGCTGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
Db      81 |AGTGGCTGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 140
Qy      541 |GAGAGTGAAGGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db      141 |GAGAGTGAAGGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 200
Qy      601 |GAGAGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db      201 |GAGAGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 260
Qy      661 |CAACAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Db      261 |CAACAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 320
Qy      721 |GATGAGATCTCTCTGCTCAAGCACTTCCCAATCCGCAAGGTCTCAAGCAAGGTCT 780
Db      321 |GATGAGATCTCTCTGCTCAAGCACTTCCCAATCCGCAAGGTCTCAAGCAAGGTCT 380
Qy      781 |GATGAGATCTCTCTGCTCAAGCACTTCCCAATCCGCAAGGTCTCAAGCAAGGTCT 840
Db      381 |GATGAGATCTCTCTGCTCAAGCACTTCCCAATCCGCAAGGTCTCAAGCAAGGTCT 440
Qy      841 |GAGGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Db      441 |GAGGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 500
Qy      901 |GAGGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
Db      501 |GAGGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 560
Qy      961 |CAATGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Db      561 |CAATGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 620
Qy      1021 |AACTACAGCTTCCCAATTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
Db      621 |AACTACAGCTTCCCAATTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 680
Qy      1081 |CCCCCGAGCTTCCCAATTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Db      681 |CCCCCGAGCTTCCCAATTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 740
Qy      1141 |ACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
Db      741 |ACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 800
Qy      1201 |CAAGGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Db      801 |CAAGGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 860
Qy      1261 |GAGGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320

```


QY 1046 GGCTCCCCCTCCGCGCCGAGAACTGCAAGCTGCCCCCGCCAGCTCCATCCACTCGA 1105
 DB 31966 GGCTCCCCCTCCGCGCCGAGAACTGCAAGCTGCCCCCGCCAGCTCCATCCACTCGA 32025
 QY 1106 TCTGCTCCCATGAGCCGGAAGCCATCCATCCCAAGCCCAAGGCGTGGAC 1165
 DB 32026 TCTGCTCCCATGAGCCGGAAGCCATCCATCCCAAGCCCAAGGCGTGGAC 32085
 QY 1166 CCGGCTCTCTTCCACTTCTTGACAGCCCAATGCGCAAGGCTTCAGAGCTCCAGCAAG 1225
 DB 32086 CCGGCTCTCTTCCACTTCTTGACAGCCCAATGCGCAAGGCTTCAGAGCTCCAGCAAG 32145
 QY 1226 CTCGCGGAGCAAGCCAGAGAGCAAGCACTTTGTAGAGTCCCAAGGCGCGAGAG 1285
 DB 32146 CTCGCGGAGCAAGCCAGAGAGCAAGCACTTTGTAGAGTCCCAAGGCGCGAGAG 32205
 QY 1286 AGTGTGGGTGTGGAGCAAGTGGAGAGAGCAAGCAAGCCCTCTGGAGAGCGCC 1345
 DB 32206 AGTGTGGGTGTGGAGCAAGTGGAGAGAGCAAGCAAGCCCTCTGGAGAGCGCC 32265
 QY 1346 ATCCCTGGGTGTGCTCCCTCCGCGCCAGCTGGAGAGCGCCCTCTCCCTCCCTA 1405
 DB 32266 ATCCCTGGGTGTGCTCCCTCCGCGCCAGCTGGAGAGCGCCCTCTCCCTCCCTA 32325
 QY 1406 GCGCCCTCTGGAGCAAGAAAGCAAGCAAGCAAGCCAGAGAGAGAGAGAGAGAGAG 1465
 DB 32326 GCGCCCTCTGGAGCAAGAAAGCAAGCAAGCAAGCCAGAGAGAGAGAGAGAGAGAG 32385
 QY 1466 GCGCTGAGAGCAAGCAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1525
 DB 32386 GCGCTGAGAGCAAGCAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 32445
 QY 1526 CTGCGCGCTGT 1585
 DB 32446 CTGCGCGCTGT 32505
 QY 1586 CACACACACCATGAAATCAACCAATACCACTTCTTACAGACATAGAGCCCTCC 1645
 DB 32506 CACACACACCATGAAATCAACCAATACCACTTCTTACAGACATAGAGCCCTCC 32565
 QY 1646 CCAAGGCGCCAGCTGCAATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1705
 DB 32566 CCAAGGCGCCAGCTGCAATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 32625
 QY 1706 CTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1765
 DB 32626 CTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 32685
 QY 1766 ATTAGCTAGCTTACATGTAGAAAGATGTAGAAACACAGAACTAACTTTATTTAT 1825
 DB 32686 ATTAGCTAGCTTACATGTAGAAAGATGTAGAAACACAGAACTAACTTTATTTAT 32745
 QY 1826 GTT 1828
 DB 32746 GTT 32748

RESULT 8
 AC007608/c 169773 bp DNA linear PRI 19-MAR-2003
 LOCUS AC007608 Homo sapiens chromosome 16 clone RP11-401P9, complete sequence.
 DEFINITION AC007608
 AC007608.7 GI:29124038
 VERSION
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 169773)
 AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los
 Alamos National Laboratory.
 TITLE Direct Submission

JOURNAL
 REFERENCE 2 (bases 1 to 169773)
 AUTHORS Bruce, D., Mund, M., Doggett, N., Munk, C., Saunders, E., Robinson, D., Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L., Bryant, J., Tesmer, J., Meincke, L., Longstre, J., White, S., Tetum, O., Campbell, J., Fawcett, J., Malbie, M., Bussod, M., Sutherland, R., McMurtry, K., Han, C. and Deaven, L.
 TITLE Direct Submission
 JOURNAL Submitted (20-MAY-1999) Center for Human Genome Studies, DOE Joint Genome Institute, Los Alamos National Laboratory, MS M888, Los Alamos, NM 87545, USA
 REFERENCE 3 (bases 1 to 169773)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (04-JAN-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 4 (bases 1 to 169773)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (11-JAN-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 5 (bases 1 to 169773)
 AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.
 TITLE Direct Submission
 JOURNAL Submitted (19-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT On Mar 19, 2003 this sequence version replaced gi:18129388.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center and Los Alamos National Laboratory
 www.sngc.stanford.edu
 Quality: Phrap Quality >=40 100% of Sequence;
 Estimated Total Number of Errors is 0.
 Location/Qualifiers
 1. 169773
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-401P9"
 ORIGIN
 Query Match 42.1%; Score 783; DB 9; Length 169773;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1046 GGCTCCCCCTCCGCGCCGAGAACTGCAAGCTGCCCCCGCCAGCTCCATCCACTCGA 1105
 DB 78524 GGCTCCCCCTCCGCGCCGAGAACTGCAAGCTGCCCCCGCCAGCTCCATCCACTCGA 78465
 QY 1106 TCTGCTCCCATGAGCCGGAAGCCATCCATCCCAAGCCCAAGGCGTGGAC 1165
 DB 78464 TCTGCTCCCATGAGCCGGAAGCCATCCATCCCAAGCCCAAGGCGTGGAC 78405
 QY 1166 CCGGCTCTCTTCCACTTCTTGACAGCCCAATGCGCAAGGCTTCAGAGCTCCAGCAAG 1225
 DB 78404 CCGGCTCTCTTCCACTTCTTGACAGCCCAATGCGCAAGGCTTCAGAGCTCCAGCAAG 78345
 QY 1226 CTCGCGGAGCAAGCCAGAGAGCAAGCACTTTGTAGAGTCCCAAGGCGCGAGAG 1285
 DB 78344 CTCGCGGAGCAAGCCAGAGAGCAAGCACTTTGTAGAGTCCCAAGGCGCGAGAG 78285
 QY 1286 AGTGTGGGTGTGGAGCAAGTGGAGAGAGCAAGCAAGCCCTCTGGAGAGCGCC 1345
 DB 78284 AGTGTGGGTGTGGAGCAAGTGGAGAGAGCAAGCAAGCCCTCTGGAGAGCGCC 78225
 QY 1346 ATCCCTGGGTGTGCTCCCTCCGCGCCAGCTGGAGAGCGCCCTCTCCCTCCCTA 1405
 DB 78224 ATCCCTGGGTGTGCTCCCTCCGCGCCAGCTGGAGAGCGCCCTCTCCCTCCCTA 78165
 QY 1406 GCGCCCTCTGGAGCAAGAAAGCAAGCAAGCAAGCCAGAGAGAGAGAGAGAGAGAG 1465

Db	78164	GCCCCCTCGGGCA	CAAGAAGCA	CAAGACCGAGCC	GAAGAGCCAG	CAGGGCTGCCG	78105
QY	1466	GGCCTGCAAG	CACCACTGAC	CTCAGGTG	CCCTGTCTTG	GGGCGGAGCACTTGC	GGAG 1525
Db	78104	GGCTGCAAG	CACCACTGAC	CTCAGGTG	CCCTGTCTTG	GGGCGGAGCACTTGC	GGAG 78045
QY	1526	CTGCCGCGCT	TTGGTGGTGTAT	GTAGAGCC	CAAGCCGGG	AGCCGGTCCAGAGCAT	GTAGCAAC 1585
Db	78044	CTGCCGCGCT	TTGGTGGTGTAT	GTAGAGCC	CAAGCCGGG	AGCCGGTCCAGAGCAT	GTAGCAAC 77985
QY	1586	CACCAACCA	CGATGAACAT	CAACCACTT	CAACGACAT	TAGAGCCCTG	CC 1645
Db	77984	CACCAACCA	CGATGAACAT	CAACCACTT	CAACGACAT	TAGAGCCCTG	CC 77925
QY	1646	CCAGGGCCCC	ACCTGCAAT	GAAAGAC	CCCAACCC	CCGACCA	CAAGGCAATTAAT 1705
Db	77924	CCAGGGCCCC	ACCTGCAAT	GAAAGAC	CCCAACCC	CCGACCA	CAAGGCAATTAAT 77865
QY	1706	CTATTAA	TTATGTTAT	TATGATGAT	TATTTAT	TATTAATAT	TATTTGTTACTCACTAAT 1765
Db	77864	CTATTAA	TTATGTTAT	TATGATGAT	TATTTAT	TATTAATAT	TATTTGTTACTCACTAAT 77805
QY	1766	ATTAGCTAG	CCCTACAT	GTAAGAT	CTATGAA	ACA	CAGACTAACTTTATTTAT 1825
Db	77804	ATTAGCTAG	CCCTACAT	GTAAGAT	CTATGAA	ACA	CAGACTAACTTTATTTAT 77745
QY	1826	GTT	1828				
Db	77744	GTT	77742				

RESULT 9					
AC007334/c	LOCUS	AC007334	170791 bp	DNA	linear PRI 13-MAR-2003
	DEFINITION	Homo sapiens chromosome 16 clone RP11-147B17, complete sequence.			
	ACCESSION	AC007334			
	VERSION	AC007334.8	GI:28933524		
	KEYWORDS	HTG.			
	SOURCE				
	ORGANISM	Homo sapiens (human)			
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE AUTHORS	TITLE	DOE Joint Genome Institute, Alamos National Laboratory. Direct Submission Unpublished	Stanford Human Genome Center and Los		
JOURNAL REFERENCE AUTHORS	TITLE	(bases 1 to 170791) Jones,D., Mundt,M., Doggett,N., Munk,C.C., Saunders,E., Robinson,D., Brice,M., Buckingham,J., Chaseen,L., Thompson,S., Goodwin,L., Bryant,J.J., Tesmer,J., Meinke,L., Longmire,J., White,S., Tatum,O., Campbell,C., Fawcett,J., Malchik,M., Bussod,M., Sutherland,R., McMurtry,K., Han,C. and Deaven,L.			
JOURNAL TITLE	TITLE	Direct Submission			
REFERENCE AUTHORS	TITLE	Submitted (21-APR-1999) Center for Human Genome Studies, DOE Joint Genome Institute, Los Alamos National Laboratory, MS M886, Los Alamos, NM 87545, USA			
JOURNAL TITLE	TITLE	(bases 1 to 170791) DOE Joint Genome Institute. Direct Submission			
REFERENCE AUTHORS	TITLE	Submitted (24-AUG-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
JOURNAL TITLE	TITLE	(bases 1 to 170791) DOE Joint Genome Institute, Alamos National Laboratory. Direct Submission	Stanford Human Genome Center and Los		
JOURNAL COMMENT	TITLE	Submitted (13-MAR-2003). DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
		On Mar 13, 2003 this sequence version replaced gi 15284265. Draft Sequence Produced by DOE Joint Genome Institute www.jgi.doe.gov			
		Finishing Completed at Stanford Human Genome Center and Los Alamos			

```

FEATURES
SOURCE
1. 170791
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="16"
   /clone="RP11-147B17"

```

ORIGIN

Query Match	42.1%	Score 783	DB 9	Length 170791
Best Local Similarity	100.0%	Pred. No. 0		
Matches 783	0	Mismatches	0	Gaps 0

1046 GGCTCCCTTCCGTGGCCAGAAGTCAGAAC TGCCCCCGACCTCCATCCACACGA 1105

DB 28069 GGCCTCCCTTACCGTAGCCAGAAAGTCAGAACATGCCCCCCTCACTCCCAAATCCCACATCCGA

1108 11108 111108 111208 111308 111408 111508 111608 111708 111808 111908 11208 112108 112208 112308 112408 112508 112608 112708 112808 112908 11308 113108 113208 113308 113408 113508 113608 113708 113808 113908 11408 114108 114208 114308 114408 114508 114608 114708 114808 114908 11508 115108 115208 115308 115408 115508 115608 115708 115808 115908 11608 116108 116208 116308 116408 116508 116608 116708 116808 116908 11708 117108 117208 117308 117408 117508 117608 117708 117808 117908 11808 118108 118208 118308 118408 118508 118608 118708 118808 118908 11908 119108 119208 119308 119408 119508 119608 119708 119808 119908 12008 120108 120208 120308 120408 120508 120608 120708 120808 120908 12108 121108 121208 121308 121408 121508 121608 121708 121808 121908 12208 122108 122208 122308 122408 122508 122608 122708 122808 122908 12308 123108 123208 123308 123408 123508 123608 123708 123808 123908 12408 124108 124208 124308 124408 124508 124608 124708 124808 124908 12508 125108 125208 125308 125408 125508 125608 125708 125808 125908 12608 126108 126208 126308 126408 126508 126608 126708 126808 126908 12708 127108 127208 127308 127408 127508 127608 127708 127808 127908 12808 128108 128208 128308 128408 128508 128608 128708 128808 128908 12908 129108 129208 129308 129408 129508 129608 129708 129808 129908 13008 130108 130208 130308 130408 130508 130608 130708 130808 130908 13108 131108 131208 131308 131408 131508 131608 131708 131808 131908 13208 132108 132208 132308 132408 132508 132608 132708 132808 132908 13308 133108 133208 133308 133408 133508 133608 133708 133808 133908 13408 134108 134208 134308 134408 134508 134608 134708 134808 134908 13508 135108 135208 135308 135408 135508 135608 135708 135808 135908 13608 136108 136208 136308 136408 136508 136608 136708 136808 136908 13708 137108 137208 137308 137408 137508 137608 137708 137808 137908 13808 138108 138208 138308 138408 138508 138608 138708 138808 138908 13908 139108 139208 139308 139408 139508 139608 139708 139808 139908 14008 140108 140208 140308 140408 140508 140608 140708 140808 140908 14108 141108 141208 141308 141408 141508 141608 141708 141808 141908 14208 142108 142208 142308 142408 142508 142608 142708 142808 142908 14308 143108 143208 143308 143408 143508 143608 143708 143808 143908 14408 144108 144208 144308 144408 144508 144608 144708 144808 144908 14508 145108 145208 145308 145408 145508 145608 145708 145808 145908 14608 146108 146208 146308 146408 146508 146608 146708 146808 146908 14708 147108 147208 147308 147408 147508 147608 147708 147808 147908 14808 148108 148208 148308 148408 148508 148608 148708 148808 148908 14908 149108 149208 149308 149408 149508 149608 149708 149808 149908 15008 150108 150208 150308 150408 150508 150608 150708 150808 150908 15108 151108 151208 151308 151408 151508 151608 151708 151808 151908 15208 152108 152208 152308 152408 152508 152608 152708 152808 152908 15308 153108 153208 153308 153408 153508 153608 153708 153808 153908 15408 154108 154208 154308 154408 154508 154608 154708 154808 154908 15508 155108 155208 155308 155408 155508 155608 155708 155808 155908 15608 156108 156208 156308 156408 156508 156608 156708 156808 156908 15708 157108 157208 157308 157408 157508 157608 157708 157808 157908 15808 158108 158208 158308 158408 158508 158608 158708 158808 158908 15908 159108 159208 159308 159408 159508 159608 159708 159808 159908 16008 160108 160208 160308 160408 160508 160608 160708 160808 160908 16108 161108 161208 161308 161408 161508 161608 161708 161808 161908 16208 162108 162208 162308 162408 162508 162608 162708 162808 162908 16308 163108 163208 163308 163408 163508 163608 163708 163808 163908 16408 164108 164208 164308 164408 164508 164608 164708 164808 164908 16508 165108 165208 165308 165408 165508 165608 165708 165808 165908 16608 166108 166208 166308 166408 166508 166608 166708 166808 166908 16708 167108 167208 167308 167408 167508 167608 167708 167808 167908 16808 168108 168208 168308 168408 168508 168608 168708 168808 168908 16908 169108 169208 169308 169408 169508 169608 169708 169808 169908 17008 170108 17020

[illegible]

27888 CTCCTGGGACCTCAGACGGAGCAAGCACTTTGTGAGTCCCCCAAGGCCCAAGGCAAG 27893

1286 AGTGTGGTGTGGCCACGTCAGAGGGGCAAGAAACAGCCCCCTCTGGACCGCC 1345

Db 27829 AGTGTGGTGTGGCCACGTGGCCAGAGGGGCAAGAAACAAGCCCCCTCTGGACCCGCC 2777

1346 ATCCCTGGGTGTCCCCCTCCGCCACCTGGCTGCCAGCCCGGCCCTCCTCCCTCCTA 1405

Db 27769 ATCCCTGGGTGTCCCCCTCCGCCCACTGGCTGCCAGCCCGCCCTCCTCCCTCCCTA 2771

1406 GCCCCCCTCGGGCACAAGAGCACAAGCACCGAGCCCAAGAGAGCCACCAAGGCTGCCG 1465

DB 2/709 G C C C C C C I C G G C A C A H A H B C H C H A B C H C L C G A B C C H A G A H A B C C A B C A H B B C I B C C B B 2/70

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

1

1545

37538 CACCAACAATGACCAATCAACCACTTCTACGACATAGAGCCCTCC 2747

1646 CCAGGGCCCACTGTCATATGAAGACCCCAACCCCGACACCAAGCATTTATT 1705

D_b 27469 CCAGGGCCCCACCCCTGCCATATGAAGGACCCCAACCCCGACACCAACAAGGATTATT 2741

1706 CTATTATTATTGTTATTATGATGATTATTGTTATTATAATTATTGTTACTCCACTAAT 1765

Db 27409 CTATTAAATTATGTTATTGTTATTGTTACTCCACTAA 27355

1766 ATTAGCTAGCCTACATGTAAGAATCTATGGAAACACAGAATCTAAATCTTTATTTATTT 1823

[illegible]

1. $\frac{1}{2}$
 2. $\frac{1}{2}$
 3. $\frac{1}{2}$
 4. $\frac{1}{2}$
 5. $\frac{1}{2}$
 6. $\frac{1}{2}$
 7. $\frac{1}{2}$
 8. $\frac{1}{2}$
 9. $\frac{1}{2}$
 10. $\frac{1}{2}$
 11. $\frac{1}{2}$
 12. $\frac{1}{2}$
 13. $\frac{1}{2}$
 14. $\frac{1}{2}$
 15. $\frac{1}{2}$
 16. $\frac{1}{2}$
 17. $\frac{1}{2}$
 18. $\frac{1}{2}$
 19. $\frac{1}{2}$
 20. $\frac{1}{2}$
 21. $\frac{1}{2}$
 22. $\frac{1}{2}$
 23. $\frac{1}{2}$
 24. $\frac{1}{2}$
 25. $\frac{1}{2}$
 26. $\frac{1}{2}$
 27. $\frac{1}{2}$
 28. $\frac{1}{2}$
 29. $\frac{1}{2}$
 30. $\frac{1}{2}$
 31. $\frac{1}{2}$
 32. $\frac{1}{2}$
 33. $\frac{1}{2}$
 34. $\frac{1}{2}$
 35. $\frac{1}{2}$
 36. $\frac{1}{2}$
 37. $\frac{1}{2}$
 38. $\frac{1}{2}$
 39. $\frac{1}{2}$
 40. $\frac{1}{2}$
 41. $\frac{1}{2}$
 42. $\frac{1}{2}$
 43. $\frac{1}{2}$
 44. $\frac{1}{2}$
 45. $\frac{1}{2}$
 46. $\frac{1}{2}$
 47. $\frac{1}{2}$
 48. $\frac{1}{2}$
 49. $\frac{1}{2}$
 50. $\frac{1}{2}$
 51. $\frac{1}{2}$
 52. $\frac{1}{2}$
 53. $\frac{1}{2}$
 54. $\frac{1}{2}$
 55. $\frac{1}{2}$
 56. $\frac{1}{2}$
 57. $\frac{1}{2}$
 58. $\frac{1}{2}$
 59. $\frac{1}{2}$
 60. $\frac{1}{2}$
 61. $\frac{1}{2}$
 62. $\frac{1}{2}$
 63. $\frac{1}{2}$
 64. $\frac{1}{2}$
 65. $\frac{1}{2}$
 66. $\frac{1}{2}$
 67. $\frac{1}{2}$
 68. $\frac{1}{2}$
 69. $\frac{1}{2}$
 70. $\frac{1}{2}$
 71. $\frac{1}{2}$
 72. $\frac{1}{2}$
 73. $\frac{1}{2}$
 74. $\frac{1}{2}$
 75. $\frac{1}{2}$
 76. $\frac{1}{2}$
 77. $\frac{1}{2}$
 78. $\frac{1}{2}$
 79. $\frac{1}{2}$
 80. $\frac{1}{2}$
 81. $\frac{1}{2}$
 82. $\frac{1}{2}$
 83. $\frac{1}{2}$
 84. $\frac{1}{2}$
 85. $\frac{1}{2}$
 86. $\frac{1}{2}$
 87. $\frac{1}{2}$
 88. $\frac{1}{2}$
 89. $\frac{1}{2}$
 90. $\frac{1}{2}$
 91. $\frac{1}{2}$
 92. $\frac{1}{2}$
 93. $\frac{1}{2}$
 94. $\frac{1}{2}$
 95. $\frac{1}{2}$
 96. $\frac{1}{2}$
 97. $\frac{1}{2}$
 98. $\frac{1}{2}$
 99. $\frac{1}{2}$
 100. $\frac{1}{2}$

```

RESULT 10
AC145238      178790 bp   DNA      linear   HTG 15-AUG-2003
LOCUS         Pan troglodytes clone CH251-568P19, WORKING DRAFT SEQUENCE, 25
DEFINITION   ordered pieces.
ACCESSION    AC145238.2 GI:33667125
VERSION      HTG, HTGS_PHASE2, HTGS_DRAFT.
KEYWORDS     Pan troglodytes (chimpanzee)
SOURCE       Pan troglodytes
ORGANISM     Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

REFERENCE
AUTHORS      Antcicelli,A., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
              Blakeley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Carliaga,K.,
              Chu,G., Coleman,B., Coleman,H., Engle,J., Granite,S., Guan,X.,
              Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,
              Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Lalic,P., Lee-Jin,S.-Q.,
              Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Mastello,C.,
              Maskeri,B., McDowell,J., Peguigian,C., Pearson,R., Portnoy,M.E.,
              Prasad,A., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K.,
              Sison,C., Stantrijop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V.,
              Vogt,J.L., Wetherby,K.D., Young,A. and Green,E.D.
              NISC Comparative Sequencing Initiative
              Unpublished
              2 (bases 1 to 178790)
              Green,E.D.
              Direct Submission
              Submitted (18-JUN-2003) NIH Intramural Sequencing Center, 8717
              Groveont Circle, Gaithersburg, MD 20877, USA
              3 (bases 1 to 178790)
              Green,E.D.
              Direct Submission
              Submitted (15-AUG-2003) NIH Intramural Sequencing Center, 8717
              Groveont Circle, Gaithersburg, MD 20877, USA
              On Aug 15, 2003 this sequence version replaced gi:31880081.

REFERENCE
AUTHORS      Center: NIH Intramural Sequencing Center
              Center code: NISC
              Web site: http://www.nisc.nih.gov
              Contact: nisc.zoo@nhgri.nih.gov
              ----- Project Information
              Center project name: eak
              Center clone name: 568P19

COMMENT
The sequence data in this record represents an 'enhanced'
version of a Phase 2 submission. Specifically, the indicated
order and orientation of each sequence contig has been
established using one or more of the following: read-pair
data from individual subclones, overlaps with neighboring
clones, alignment with available reference sequence (e.g.,
human), and/or confirmation by PCR testing. In addition,
the sequence assembly is based on at least 8X average
coverage in Q20 bases and has been reviewed to rule out
gross misassemblies, the low-quality ends of sequence
contigs have been trimmed away, and each base is associated
with a Phrap-derived quality score.
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171494 bases at least Q40
Consensus quality: 174120 bases at least Q30
Consensus quality: 175495 bases at least Q20
Insert size: 149000; agarose-efp
Insert coverage: 176390; sum-of-contigs
Quality coverage: 9.22x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces

```

```

* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1
2604: contig of 2604 bp in length
2605      2704: gap of unknown length
2705      7790: contig of 5086 bp in length
7791      7890: gap of unknown length
7891      12833: contig of 4943 bp in length
12834      12933: gap of unknown length
12934      14664: contig of 1731 bp in length
14665      14764: gap of unknown length
14765      32237: contig of 17473 bp in length
32238      32337: gap of unknown length
32338      35356: contig of 3019 bp in length
35357      35456: gap of unknown length
35457      39292: contig of 3836 bp in length
39293      39392: gap of unknown length
39393      41210: contig of 1818 bp in length
41211      43762: contig of 2452 bp in length
43763      43862: gap of unknown length
43863      46766: contig of 2904 bp in length
46767      46866: gap of unknown length
46867      53405: contig of 6533 bp in length
53406      53505: gap of unknown length
53506      55334: contig of 1829 bp in length
55335      55434: gap of unknown length
55435      69095: contig of 13661 bp in length
69096      69195: gap of unknown length
69196      78811: contig of 9616 bp in length
78812      78911: gap of unknown length
78912      89715: contig of 10804 bp in length
89716      89815: gap of unknown length
89816      97259: contig of 7444 bp in length
97260      97359: gap of unknown length
97360      99238: contig of 1879 bp in length
99239      99338: gap of unknown length
99340      101324: contig of 1986 bp in length
101325      101424: gap of unknown length
101425      112068: contig of 10664 bp in length
112069      112168: gap of unknown length
112169      118779: contig of 6511 bp in length
118780      118779: gap of unknown length
118781      120780: contig of 1901 bp in length
120781      120780: gap of unknown length
120782      135732: contig of 14952 bp in length
135733      135832: gap of unknown length
135833      138723: contig of 2891 bp in length
138724      138823: gap of unknown length
138824      163652: contig of 24829 bp in length
163653      163752: gap of unknown length
163753      178790: contig of 15038 bp in length.

FEATURES
Location/Qualifiers
1..178790
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="CH251-568P19"
/clone_1b="CH251"
1..90615
/clone="clone overlaps with GenBank Accession Number
AC145175 clone CH251-32B1 (center project name eaj)"
1..2604
/clone="assembly_fragment"
clone_end=896
vector_side="left"
2705..7790
/clone="assembly_fragment"
7891..12833
/clone="assembly_fragment"

```

```
misc_feature      12934..14664
                    /note="assembly_fragment"
misc_feature      14765..32237
                    /note="assembly_fragment"
misc_feature      32338..35356
                    /note="assembly_fragment"
misc_feature      35457..39292
                    /note="assembly_fragment"
misc_feature      39393..41210
                    /note="assembly_fragment"
misc_feature      41311..43762
                    /note="assembly_fragment"
misc_feature      43863..46766
                    /note="assembly_fragment"
misc_feature      46867..53405
                    /note="assembly_fragment"
misc_feature      53506..55334
                    /note="assembly_fragment"
misc_feature      55435..69095
                    /note="assembly_fragment"
misc_feature      69196..78811
                    /note="assembly_fragment"
misc_feature      78912..89715
                    /note="assembly_fragment"
misc_feature      89816..97259
                    /note="assembly_fragment"
misc_feature      97360..99238
                    /note="assembly_fragment"
misc_feature      99339..101324
                    /note="assembly_fragment"
misc_feature      101425..112068
                    /note="assembly_fragment"
misc_feature      112169..118679
                    /note="assembly_fragment"
misc_feature      118780..120680
                    /note="assembly_fragment"
misc_feature      120781..135732
                    /note="assembly_fragment"
misc_feature      135833..138723
                    /note="assembly_fragment"
misc_feature      138824..178790
                    /note="clone overlaps with GenBank Accession Number
                    AC145174 clone CH251-260A1 (center project name es1)"
misc_feature      138824..163652
                    /note="assembly_fragment"
misc_feature      163753..178790
                    /note="assembly_fragment
                    clone end:T7
                    vector_side:right"

ORIGIN
Query Match      28.4%; Score 528; DB 2; Length 178790;
Best Local Similarity 99.4%; Pred. No. 5.9e-280;
Matches 778; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1046 GGCCTCCCTTCCGCGCCGAGTCAAGTGCCTCCGCACTCCATCCACTCGA 1105
DB      65319 GGCCTCCCTTCCGCGCCGAGTCAAGTGCCTCCGCACTCCATCCACTCGA 65378
QY      1106 TCTGCTCCCATGAGCGGAAGCATCCATCCACACCGAAGCCCAAGCGGTGAC 1165
DB      65379 TCTGCTCCCATGAGCGGAAGCATCCATCCACACCGAAGCCCAAGCGGTGAC 65438
QY      1166 CCGGCTCTTTCATTCTTTCAGACCCCAATGCGCAAGTCTCAGAGCTTCAGCAACG 1225
DB      65439 CCGGCTCTTTCATTCTTTCAGACCCCAATGCGCAAGTCTCAGAGCTTCAGCAACG 65498
QY      1226 CTCCTGGGACCCGAGACGGAGCACTTTGTGAGTCCCGCAAGCCCAAGGCAAG 1285
DB      65499 CTCCTGGGACCCGAGACGGAGCACTTTGTGAGTCCCGCAAGGCAAGGCAAG 65558
QY      1286 AGTGTGGGTGTGGCCAGTGGCCAGAGGGGCAAGAAACAAGCCCTCTGGGACCGCC 1345
```

```
DB      65559 AGTGTGGGTGTGGCCAGTGGCCAGAGGGGCAAGAAACAAGCCCTCTGGGACCGCC 65618
QY      1346 ATCCCTGGGTGTGTCCTCCCGCCACCTGGCTGCGAGCCCGGCTCTCCCTCCCTA 1405
DB      65619 ATCCCTGGGTGTGTCCTCCCGCCACCTGGCTGCGAGCCCGGCTCTCCCTCCCTA 65678
QY      1406 GCGCCCTCGGGGCAAGAAGCAACAAGCAGACCGAGCAAGAGCCAGCGGCTCCG 1465
DB      65679 GCGCCCTCGGGGCAAGAAGCAACAAGCAGACCGAGCAAGAGCCAGCGGCTCCG 65738
QY      1466 GGCCTGAGGACACACTGCGCTCAGGTGCGCTGTCTCTGGGCGGAGACCTGCGGAG 1525
DB      65739 GGCCTGAGGACACACTGCGCTCAGGTGCGCTGTCTCTGGGCGGAGACCTGCGGAG 65798
QY      1526 CTGCCCCCTTGGTGTGTATGAGACCGAGCGGGGAGCCGGTCCAGAGATGAGCAC 1585
DB      65799 CTGCCCCCTTGGTGTGTATGAGACCGAGCGGGGAGCCGGTCCAGAGATGAGCAC 65858
QY      1586 CACCAACCATGAAATGACACCACTTACCACTTACCAAGATAGAGCCCTCC 1645
DB      65859 CACCAACCATGAAATGACACCACTTACCACTTACCAAGATAGAGCCCTCC 65918
QY      1646 CAGAGGCCCAACCTGCGCATATGAGAGACCCACCCCGACACCAAGCATTAATT 1705
DB      65919 CAGAGGCCCAACCTGCGCATATGAGAGACCCACCCCGACACCAAGCATTAATT 65978
QY      1706 CTATTATTAATTGTTATTAATGATTAATTATTAATTATTAATTACTGCACTAAT 1765
DB      65979 CTATTATTAATTGTTATTAATGATTAATTATTAATTATTAATTACTGCACTAAT 66038
QY      1766 ATTAGCTAGCTCATATGAGAAATCTATGAAACACAGAACTTTATTTATAT 1825
DB      66039 ATTAGCTAGCTCATATGAGAAATCTATGAGAAACACAGAACTTTATTTATAT 66098
QY      1826 GTT 1828
DB      66099 GTT 66101

RESULT 11
AC145175      207486 bp      DNA      linear      HTG 15-AUG-2003
LOCUS      Pan troglodytes clone CH251-32B1, WORKING DRAFT SEQUENCE, 11
DEFINITION      ordered pieces.
AC145175.2      GI:33667124
VERSION      HTG; HTGS_PHASE2; HTGS_DRAFT.
KEYWORDS      Pan troglodytes (chimpanzee)
SOURCE      Pan troglodytes
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE      1 (bases 1 to 207486)
AUTHORS      Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
      Blakesley,R.W., Boufard,G.G., Brinkley,C., Brooks,S., Carliaga,K.,
      Chu,G., Coleman,B., Coleman,H., Engle,J., Granite,S., Guan,X.,
      Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-J., Hu,P.,
      Harle,B., Idol,J.R., Karlins,E., Khong,P., Latic,P., Lee-Jin,S.-O.,
      Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masello,C.,
      Maseri,B., McDowell,J., Paguirigan,C., Pearson,R., Portnoy,M.E.,
      Prasad,A., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K.,
      Sison,C., Stanciripop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V.,
      Vogt,J.L., Wetherby,K.D., Young,A. and Green,E.D.
      NISC Comparative Sequencing Initiative
      Unpublished
      2 (bases 1 to 207486)
REFERENCE      Green,E.D.
AUTHORS      Direct Submission
TITLE      Submitted (12-JUN-2003) NIH Intramural Sequencing Center, 8717
      Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE      3 (bases 1 to 207486)
AUTHORS      Green,E.D.
TITLE      Direct Submission
JOURNAL      Submitted (15-AUG-2003) NIH Intramural Sequencing Center, 8717
```


Qy	1646	CCAGGCCCCACCCCGCCATATGAAGGACCCACCCCGACACCAAGGCATTATATT	1705
Db	181067	CCAGGCCCCACCCCGCCATATGAAGGACCCACCCCGACACCAAGGCATTATATT	181126
Qy	1706	CTATTAAATTAATGTATATATGATATATTTGTATTAATTAATTAATTTGTTACTCACTAAT	1765
Db	181127	CTATTAAATTAATGTATATATGATATATTTGTATTAATTAATTAATTTGTTACTCACTAAT	181186
Qy	1766	ATTAGCTAGCCTACATGAGAAAGATCTATGGAACAACAGACTAATCTTTATTTATAT	1825
Db	181187	ATTAGCTAGCCTACATGAGAAAGATCTATGGAACAACAGACTAATCTTTATTTATAT	181246
Qy	1826	GTT 1828	
Db	181247	GTT 181249	
RESULT 12			
LOCUS	AX557538	1950 bp	DNA
DEFINITION	Sequence 126 from Patent WO0248370.	linear	PAT 27-NOV-2002
ACCESSION	AX557538		
VERSION	AX557538.1	GI:25900445	
KEYWORDS			
ORGANISM	Homo sapiens (human)		
SOURCE	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 Sun, Y., Recipon, H., Ghosh, M. G. and Liu, C. Compositions and methods relating to colon specific genes and proteins		
JOURNAL	Patent: WO 0248370-A 126 20-JUN-2002;		
FEATURES	Diadexus, Inc. (US)		
SOURCE	location/Qualifiers		
	1..1950		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query_Match	7.4%;	Score 137;	DB 6; Length 1950;
Best Local Similarity	100.0%;	Pred No. 8.3e-64;	
Matches 137;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Qy	281	GGTGAACGCTTCGCCGCTGAGCGCTGCCTGCGCTCGAAGGCGATCGAGAGTGATCGG	340
Db	838	GGTGAACGCTTCGCCGCTGAGCGCTGCCTGCGCTCGAAGGCGATCGAGAGTGATCGG	897
Qy	341	AGACAGCGCTGCCCGGGCGGTGTCTGGGACCCCGACAGCTGGCGTTGGGGGACCAT	400
Db	898	AGACAGCGCTGCCCGGGCGGTGTCTGGGACCCCGACAGCTGGCGTTGGGGGACCAT	957
Qy	401	GGCCGAGCACCCGGGA 417	
Db	958	GGCCGAGCACCCGGGA 974	
RESULT 13			
LOCUS	CQ741468	2340 bp	DNA
DEFINITION	Sequence 27402 from Patent WO02068579.	linear	PAT 03-FEB-2004
ACCESSION	CQ741468		
VERSION	CQ741468.1	GI:42349559	
KEYWORDS			
ORGANISM	Homo sapiens (human)		
SOURCE	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 Venter, C. J., Adams, M. C., Li, P. W. and Myers, E. W. Kite, such as nucleic acid arrays, comprising a majority of humaneons or transcripts, for detecting expression and other uses		

JOURNAL	threecf
PATENT	Patent: WO 02068579-A 27402 06-SEP-2002;
FEATURES	PG Corporation (NY) (US)
SOURCE	Location/Qualifiers
	1. .2340
	/organism="Homo sapiens"
	/mol_type="unassigned DNA"
	/db_xref="taxon:9606"
ORIGIN	
Query Match	7.3%; Score 136; DB 6; Length 2340;
Best Local Similarity	100.0%; Pred. No. 3e-63;
Matches 136; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	280 AGGTGACAGCTTCCCGCTGAGCGCTGCTGAGGAGCATCGAGAGTGGATCGG 339
DB	1524 AGGTGACAGCTTCCCGCTGAGCGCTGCTGAGGAGCATCGAGAGTGGATCGG 1583
OY	340 GAGA CAGCGCTGCCCGGCGGTGTCTCGGAGACCCGACAGCTGCGTGGCGGACCAT 399
DB	1584 GAGACAGCGCTGCCCGGCGGTGTCTCGGAGACCCGACAGCTGCGTGGCGGACCAT 1643
OY	400 AGCGCGAAGCACCCGG 415
DB	1644 AGCGCGAAGCACCCGG 1659
RESULT 14	
LOCUS	AC145003 194869 bp DNA linear HTG 03-JUL-2003
DEFINITION	Papio anubis clone RP41-205A15, WORKING DRAFT SEQUENCE, 9 ordered
FEATURES	pieces.
ACCESSION	AC145003
VERSION	AC145003.2 GI:32441299
KEYWORDS	HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE	Papio anubis (olive baboon)
ORGANISM	Papio anubis
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
	Cercopithecinae; Papio.
REFERENCE	1 (bases 1 to 194869)
AUTHORS	Antonellis A., Ayele K., Beckstrom-stenberg S.M., Benjamin B.,
	Blakesley R.W., Bouffard G.G., Brinkley C., Brooks S., Carliaga K.,
	Chu G., Coleman B., Coleman H., Engle J., Granite S., Guan X.,
	Gupta J., Haghighi P., Han J., Hansen N., Ho S.-L., Hu P.,
	Hupfe B., Idol J.R., Karling B., Kwon P., Latic P., Lee-Lin S.-Q.,
	Legaspi R., Maduro Q.L., Maduro V.B., Margulies E.H., Mastello C.,
	Maskei B., McDowell J., Pasquigan C., Pearson R., Portnoy M.E.,
	Prasad A., Reddik-Bugne N., Schandler K., Scheller M.G., Shah K.,
	Sison C., Stenitrop S., Thomas J.W., Thomas P.J., Taipouri V.,
	Voet J.L., Weetherby K.D., Wiggins L., Young A. and Green E.D.
	NISC Comparative Sequencing Initiative
TITLE	Unpublished
JOURNAL	2 (bases 1 to 194869)
REFERENCE	Green E.D.
AUTHORS	Direct Submission
TITLE	Submitted (30-MAY-2003) NIH Intramural Sequencing Center, 8717
JOURNAL	Groveton Circle, Gaithersburg, MD 20877, USA
REFERENCE	3 (bases 1 to 194869)
AUTHORS	Green E.D.
TITLE	Direct Submission
JOURNAL	Submitted (03-JUL-2003) NIH Intramural Sequencing Center, 8717
COMMENT	Groveton Circle, Gaithersburg, MD 20877, USA
	On Jul 3, 2003 this sequence version replaced gi:31193878.
	----- Genome Center
	Center: NIH Intramural Sequencing Center
	Center code: NISC
	Web site: http://www.nisc.nih.gov
	Contact: nisc.zoo@nih.gov
	----- Project Information
	Center project name: dtr
	Center clone name: 205A15

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics -----

Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 193209 bases at least Q40
 Consensus quality: 193760 bases at least Q30
 Consensus quality: 193984 bases at least Q20
 Insert size: 180000; agarose-fp
 Insert size: 194069; sum-of-contigs
 Quality coverage: 14.34x in Q20 bases; agarose-fp
 Quality coverage: 13.30x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

```

1      7609      7608: contig of 7608 bp in length
*      7709      10151: contig of 2443 bp in length
*      10152      10251: gap of unknown length
*      10252      44876: contig of 34625 bp in length
*      44877      44976: gap of unknown length
*      44977      76344: contig of 31368 bp in length
*      76345      76444: gap of unknown length
*      76445      78547: contig of 2103 bp in length
*      78548      78647: gap of unknown length
*      78648      145819: contig of 67172 bp in length
*      145820      145919: gap of unknown length
*      145920      163253: contig of 17334 bp in length
*      163254      163353: gap of unknown length
*      163354      166218: contig of 2765 bp in length
*      166219      166218: gap of unknown length
*      166219      194869: contig of 28651 bp in length.

```

FEATURES

```

source
  1. 194869
    /organism="Papio anubis"
    /mol_type="genomic DNA"
    /db_xref="taxon:9555"
    /clone="RP41-205A15"
    /clone_1b="RP41"
  1. 31748
    /note="clone overlaps with GenBank Accession Number AC145043 clone RP41-2B13 (center project name dtg)"
  1. 7608
    /note="assembly_fragment"
    clone_end:17
    vector_side:left
  7709..10151
    /note="assembly_fragment"
  10252..44876
    /note="assembly_fragment"
  44977..76344
    /note="assembly_fragment"
  76445..78547
    /note="assembly_fragment"
  78648..145819
    /note="assembly_fragment"

```

```

/misc_feature      /note="assembly_fragment"
145920..163253
/misc_feature      /note="assembly_fragment"
163354..166118
/misc_feature      /note="assembly_fragment"
166219..194869
/misc_feature      /note="assembly_fragment"
clone_end:196
vector_side:right

```

ORIGIN

Query Match 6.7%; Score 124; DB 2; Length 194869;
 Best Local Similarity 99.4%; Pred. No. 1.7e-56;
 Matches 174; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy      1171 CTCCTTCACATCTCTTACACCCCAATCCGCAAGTCTCAGAGTCAGCAAGCGCTCCG 1230
Db      16651 CTCCTTCACATCTCTTACACCCCAATCCGCAAGTCTCAGAGTCAGCAAGCGCTCCG 16710
Qy      1231 GGGCACCACGAGCGGAGCAAGCACTTGTGAGTCTCCCAAGGCCAGGCAAGACTGT 1290
Db      16711 GGGCACCACGAGCGGAGCAAGCACTTGTGAGTCTCCCAAGGCCAGGCAAGACTGT 16770
Qy      1291 GGGGTCGGCGCAGTGGCGCAGAGGGCGCAAGAACAGCCCTCTGAGACCCGCC 1345
Db      16771 GGGGTCGGCGCAGTGGCGCAGAGGGCGCAAGAACAGCCCTCTGAGACCCGCC 16825

```

RESULT 15

AC145043 198189 bp DNA linear HTG 26-JUN-2003
 LOCUS Papio anubis clone RP41-2B13, WORKING DRAFT SEQUENCE, 4 ordered pieces.

ACCESSION AC145043
 VERSION AC145043.2 GI:32261339
 KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT.
 SOURCE Papio anubis (olive baboon)
 ORGANISM Papio anubis

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopitheciinae; Papio.
 1 (bases 1 to 198189)
 Blakesley, A., Ayale, K., Beckstrom-Stenberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Carriaga, K., Chu, G., Coleman, B., Coleman, H., Engle, J., Granite, S., Guan, X., Gupta, J., Haghighi, P., Han, D., Hansen, N., Ho, S.-L., Hu, P., Hurle, B., Idol, J.R., Karlins, E., Kwong, P., Laric, P., Lee-Lin, S.-O., Legaspi, R., McDowell, J., Paguitigan, C., Pearson, R., Portnoy, M.E., Prasad, A., Reddix-Dugue, N., Schandier, K., Schneider, M.G., Shah, K., Sison, C., Stantridop, S., Thomas, J.W., Thomas, P.D., Tsipouris, V., Vogt, J.L., Wetherby, K.D., Wiggins, L., Young, A. and Green, E.D.
 NISC Comparative Sequencing Initiative
 Unpublished

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: http://www.nisc.nih.gov
 Contact: nisc.zoo@nih.gov
 Project Information
 Center project name: dtg
 Center clone name: 002B13

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics -----

Sequencing vector: plasmid, n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 197710 bases at least Q40
Consensus quality: 197833 bases at least Q30
Consensus quality: 197878 bases at least Q20
Insert size: 193000; agarose-fp
Insert size: 197889; sum-of-contigs
Quality coverage: 11.84x in Q20 bases; agarose-fp
Quality coverage: 11.55x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 20663: contig of 20663 bp in length
* 20664 20763: gap of unknown length
* 20764 173988: contig of 153225 bp in length
* 173989 174088: gap of unknown length
* 174089 176651: contig of 2563 bp in length
* 176652 176751: gap of unknown length
* 176752 198189: contig of 21438 bp in length.
Location/Qualifiers
1. .198189
/organism="Papio anubis"
/mol_type="genomic DNA"
/db_xref="taxon:9555"
/clone="RP41-2E13"
/clone_id="RP41"
1. .20663
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
20764. .173988
/note="assembly_fragment"
166409. .198189
/note="clone overlaps with GenBank Accession Number
AC145003 clone RP41-205A15 (center project name drr)"
174089. .176651
/note="assembly_fragment"
176752. .198189
/note="assembly_fragment"
missing T7 clone end on 3' end of insert"

FEATURES
source

misc_feature
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
20764. .173988
/note="assembly_fragment"
166409. .198189
/note="clone overlaps with GenBank Accession Number
AC145003 clone RP41-205A15 (center project name drr)"
174089. .176651
/note="assembly_fragment"
176752. .198189
/note="assembly_fragment"
missing T7 clone end on 3' end of insert"

ORIGIN

Query Match 6.7%; Score 124; DB 2; Length 198189;
Best Local Similarity 99.4%; Pred.No. 1.7e-56;
Matches 174; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1171 CTCCTTCACATCTCTTGACATCCCAATGCGCAAGTCTCAGAGTCCAGCAACGGCTCCG 1230
183091 CTCCTTCACATCTCTTGACATCCCAATGCGCAAGTCTCAGAGTCCAGCAACGGCTCCG 183150
1231 GGGACCCAGGACGGAGCAAGCACTTGAGGTCCCGCAAGGCCAGGCAAGAGTGT 1290

Db 183151 GGGACCCAGGACGGAGCAAGCACTTGAGGTCCCGCAAGGCCAGGCAAGAGCGT 183210
QY 1291 GGGTGTGGGCCACGTGCGCCAGAGGGGCAAGAAACAAGCCCCCTGTGGACCCGCC 1345
Db 183211 GGGTGTGGGCCACGTGCGCCAGAGGGGCAAGAAACAAGCCCCCTGTGGACCCGCC 183265

Search completed: December 30, 2004, 00:53:14
Job time : 12710 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2004, 16:38:42 ; Search time 1269 Seconds
(without alignments)
7690.050 Million cell updates/sec

Title: US-09-993-966-5
Sequence: 1 gaattcgccctctactacg.....aaaaaaaaaagcgccgc 1859

Scoring table: ~~OLIGOBLAST~~
Gapop 60.0 , Gapext 60.0

Searched: 413486 seqs, 2624710521 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_23Sep04:*
1: geneseqn19808:*
2: geneseqn19908:*
3: geneseqn20008:*
4: geneseqn20018:*
5: geneseqn20018:*
6: geneseqn20028:*
7: geneseqn20028:*
8: geneseqn20038:*
9: geneseqn20038:*
10: geneseqn20038:*
11: geneseqn20038:*
12: geneseqn20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1859	100.0	1859	10	AAL51511 Human Nkd
2	1336	71.9	1438	3	AAAG3925 DNA encod
3	783	42.1	2379	4	AAL03214 Human rep
4	783	42.1	2379	4	AAL03216 Human rep
5	590	31.7	590	10	AAL51525 Human Nkd
6	405	21.8	1561	5	AAAS2082 DNA encod
7	377	20.3	427	5	ACH21474 Human adu
8	218	11.7	1743	5	AA578752 DNA encod
9	218	11.7	1743	5	AA592081 DNA encod
10	218	11.7	1743	10	ADG31966 Human nov
11	215	11.6	215	10	AAL51526 Human Nkd
12	148	8.0	148	10	AAL51522 Human Nkd
13	137	7.4	598	12	ACH74795 Human gen
14	137	7.4	1950	6	ABO73820 Human col
15	134	7.2	134	10	AAL51518 Human Nkd
16	133	7.2	133	12	ACH88495 Human gen
17	128	6.9	128	10	AAL51524 Human Nkd
18	120	6.5	120	10	AAL51514 Human Nkd
19	108	5.8	639	8	ACD05637 cDNA enco
20	107	5.8	107	10	AAL51520 Human Nkd
21	96	5.2	96	10	AAL51521 Human Nkd

ALIGNMENTS

22	88	4.7	554	12	ACH74420	ACH74420 Human gen
23	86	4.6	179	12	ACH88120	ACH88120 Human gen
24	85	4.6	85	10	AAL51523	AAL51523 Human Nkd
25	67	3.6	67	10	AAL51519	AAL51519 Human Nkd
26	58	3.1	670	6	ABV73487	ABV73487 Human cyt
27	58	3.1	1264	12	ADO24765	ADO24765 Dog TRAIL
28	57	3.1	417	6	ABV73496	ABV73496 Human cyt
29	57	3.1	510	6	ABV73485	ABV73485 Human cyt
30	56	3.0	636	6	ABV73493	ABV73493 Human cyt
31	56	3.0	651	6	ABV73476	ABV73476 Human cyt
32	56	3.0	686	6	ABV73486	ABV73486 Human cyt
33	56	3.0	707	6	ABX04893	ABX04893 Conus sp
34	55	3.0	547	6	ABV73478	ABV73478 Human cyt
35	55	3.0	1245	6	ABV73473	ABV73473 Human cyt
36	54	2.9	1874	10	AD82682	AD82682 Terpenoid
37	54	2.9	1874	10	AD82678	AD82678 Terpenoid
38	54	2.9	2605	10	AD82704	AD82704 Terpenoid
39	53	2.9	1434	6	ABA92108	ABA92108 Soybean p
40	53	2.9	2755	10	AD86813	AD86813 Petunia p
41	53	2.9	3281	10	ADG32558	ADG32558 Murine TR
42	52	2.8	577	6	ABV73482	ABV73482 Human cyt
43	52	2.8	1124	6	AAL46702	AAL46702 B tropica
44	52	2.8	1139	6	AAL46704	AAL46704 B tropica
45	52	2.8	1168	6	AAL46703	AAL46703 B tropica

RESULT 1
AAL51511
ID AAL51511 standard; DNA; 1859 BP.
XX
AC AAL51511;
XX
DT 24-APR-2003 (first entry)
XX
DE Human Nkd (hNkd) protein coding sequence.
XX
KW Human; gene; ds; gene therapy; Nkd; hNkd; wnt signalling pathway; cancer;
KW colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 224..1636
FT /*tag= a
FT /product= "Human Nkd protein"
XX
FN WO200292832-A2.
XX
PD 21-NOV-2002.
XX
PF 27-NOV-2001; 2001WO-US044092.
XX
PR 27-NOV-2000; 2000US-0252884P.
PR 16-MAY-2001; 2001US-0291109P.
PR 01-OCT-2001; 2001US-0325571P.
XX
PA (CHIR) CHIRON CORP.
XX
FI Rohan M, Chan V, Yan D;
XX
DR WPI; 2003-129303/12.
DR P-PSDB; AAO16389.
XX
PT New human and non-human primate homologues of Nkd protein, and Nkd genes,
PT useful for treating cancer involving aberrant wnt signalling, e.g. colon
PT cancer, head and neck cancer, ovarian cancer, or breast cancer.
XX
PS Claim 4; Fig 3; 99pp; English.
XX
CC The invention comprises the amino acid and coding sequence of the human

CC Nkd protein (hmkd) - a regulator of Wnt signalling pathways. The hmkd DNA
CC and protein sequences are useful for the treatment of cancer involving
CC aberrant Wnt signalling (e.g. colon cancer, head and neck cancer, ovarian
CC cancer and breast cancer). The present DNA sequence encodes the human Nkd
CC protein of the invention
XX

Sequence 1859 BP; 434 A; 587 C; 551 G; 287 T; 0 U; 0 Other;

Query Match 100.0%; Score 1859; DB 10; Length 1859;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 1859; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCGCGCTTCAATACGACTCACTATAGGCGAAGAGTGAACAAGCAAGTAAAG 60
DB 1 GAATTCGCGCTTCAATACGACTCACTATAGGCGAAGAGTGAACAAGCAAGTAAAG 60
QY 61 CGGGAGATCGGGCCGCGCGAGCGCGGAGAGCGCTCCGCGCGCGCTCGGCTCCG 120
DB 61 CGGGAGATCGGGCCGCGCGAGCGCGGAGAGCGCTCCGCGCGCGCTCGGCTCCG 120
QY 121 CTCGGCTCGGGGCTGCTCTGGGAGAGAGAGTCAAGAGAGTCCAGGCTCCGCGG 180
DB 121 CTCGGCTCGGGGCTGCTCTGGGAGAGAGAGTCAAGAGAGTCCAGGCTCCGCGG 180
QY 181 GGGCGATAGGCTTAGGAGACGCTCCGCGCGCGAGCCAGCATGGGGAAACTTCACTC 240
DB 181 GGGCGATAGGCTTAGGAGACGCTCCGCGCGCGAGCCAGCATGGGGAAACTTCACTC 240
QY 241 CAAGCCGCGCGCGCTGTGCAAGCGCAGAGAGAGCCGGAAGGTGACAGCTTCCGCTGAG 300
DB 241 CAAGCCGCGCGCGCTGTGCAAGCGCAGAGAGAGCCGGAAGGTGACAGCTTCCGCTGAG 300
QY 301 CGCTGCTGGGCTCGGAGAGGCAATCGAGAGTGAATGGGAGAGACAGGCTCCGCGG 360
DB 301 CGCTGCTGGGCTCGGAGAGGCAATCGAGAGTGAATGGGAGAGACAGGCTCCGCGG 360
QY 361 TGTCTCGGAGACCCGAGACGCTGCGGAGCACTATAGGCTCGGAGACT 420
DB 361 TGTCTCGGAGACCCGAGACGCTGCGGAGCACTATAGGCTCGGAGACT 420
QY 421 CGTGGGCGACGTGTTGAAGACAGCTCAAGGAGAAAGAGAGAGCACTTTCGCTGGA 480
DB 421 CGTGGGCGACGTGTTGAAGACAGCTCAAGGAGAAAGAGAGAGCACTTTCGCTGGA 480
QY 481 AGTGGCCCTGCTCTGAGAGAGTGAAGAGTGAATGGGAGAGAGAGAGAGAGAG 540
DB 481 AGTGGCCCTGCTCTGAGAGAGTGAAGAGTGAATGGGAGAGAGAGAGAGAGAG 540
QY 541 GAGAGTGAAGCACTGCGCAGAGCTCAAGAGAGAGTGAAGTGAAGAGTCAAGT 600
DB 541 GAGAGTGAAGCACTGCGCAGAGCTCAAGAGAGAGTGAAGTGAAGAGTCAAGT 600
QY 601 CGAGGTGTCAATGAGAGAGAGACAGCCGCGAGAGTGAAGTCACTTCACTTGA 660
DB 601 CGAGGTGTCAATGAGAGAGAGACAGCCGCGAGAGTGAAGTCACTTCACTTGA 660
QY 661 CAAGAAGCGCAAGGTCAACCGAGAGAGACATCAACAGTGTGCAACCACTTGAAGT 720
DB 661 CAAGAAGCGCAAGGTCAACCGAGAGAGACATCAACAGTGTGCAACCACTTGAAGT 720
QY 721 GGTGGAATCTCTGTCAACCACTCCCAACATCCAGCAAGATGCTGGGGTAAAGCTCAC 780
DB 721 GGTGGAATCTCTGTCAACCACTCCCAACATCCAGCAAGATGCTGGGGTAAAGCTCAC 780
QY 781 CGTGGCCCGAGAGAGAGAGCAAGAGAGAGTCTTGTCAATCAAGCTGACCTGCA 840
DB 781 CGTGGCCCGAGAGAGAGAGCAAGAGAGAGTCTTGTCAATCAAGCTGACCTGCA 840
QY 841 GAGGCGAAGGCGCCGAGAGAGACCAAGCCCACTGAGAGCTTGCGAGCTGGAGAGAA 900
DB 841 GAGGCGAAGGCGCCGAGAGAGACCAAGCCCACTGAGAGCTTGCGAGCTGGAGAGAA 900
QY 901 GCAAGGAGCCCGCTCAGGTTCCAGGGTGAACGCGCTGAGAGAGTCTGCTGCTACA 960

DB 901 GCAAGGAGCCCGCTCAGGTTCCAGGGTGAACGCGCTGAGAGAGTCTGCTGCTACA 960
QY 961 CCAATGCGTAGATGAGAAATCAAGAGAGAGAAACCACTACTTATGATTCGCCGGAGTGA 1020
DB 961 CCAATGCGTAGATGAGAAATCAAGAGAGAGAAACCACTACTTATGATTCGCCGGAGTGA 1020
QY 1021 AAATCAAGCTCCCAATTTGGGCTGGCTCCCTCCGCGGAGAGTCAAGTCAAGTGGC 1080
DB 1021 AAATCAAGCTCCCAATTTGGGCTGGCTCCCTCCGCGGAGAGTCAAGTCAAGTGGC 1080
QY 1081 CCCCAGACCTCCAAATCCCACTGATCTCGCTCCCATGAGCCGGAAGCATTCACATCCC 1140
DB 1081 CCCCAGACCTCCAAATCCCACTGATCTCGCTCCCATGAGCCGGAAGCATTCACATCCC 1140
QY 1141 ACAACGAAGCCCAAGGCGTGAACCCGCGCTCTCTTCACTTCTTGAACATCCATCCG 1200
DB 1141 ACAACGAAGCCCAAGGCGTGAACCCGCGCTCTCTTCACTTCTTGAACATCCATCCG 1200
QY 1201 CAAGTCTCAGAGCTCCAGCAAGGCTCCGCGGACCCAGAGAGGAGCACTTGT 1260
DB 1201 CAAGTCTCAGAGCTCCAGCAAGGCTCCGCGGACCCAGAGAGGAGCACTTGT 1260
QY 1261 GAGGTCCCAAGGCGCAAGGCGCAAGTGTGAGTGTGAGCAGAGTGGCCAGAGGCGCAG 1320
DB 1261 GAGGTCCCAAGGCGCAAGGCGCAAGTGTGAGTGTGAGCAGAGTGGCCAGAGGCGCAG 1320
QY 1321 AAACAAGCCCTCTGTGGAGACCCGCACTCTCGGTGTCTCCCTCCGCGCACTGCTGTC 1380
DB 1321 AAACAAGCCCTCTGTGGAGACCCGCACTCTCGGTGTCTCCCTCCGCGCACTGCTGTC 1380
QY 1381 CAGCGCGGCTCTCTCCCTCCCTAGCGCCCTCGGAGCAAGAGAGCAAGCAGCAGC 1440
DB 1381 CAGCGCGGCTCTCTCCCTCCCTAGCGCCCTCGGAGCAAGAGAGCAAGCAGCAGC 1440
QY 1441 CAAGAGAGCGCAGAGGCTGCGGAGCTGCAAGCACTGAGCTCAGAGTGGCTGT 1500
DB 1441 CAAGAGAGCGCAGAGGCTGCGGAGCTGCGGAGCTGCAAGCACTGAGCTCAGAGTGGCTGT 1500
QY 1501 CTTGGGCGGAGACCTGCGGAGCTGCGGCTTGTGTGTATGAGAGCCAGGCGG 1560
DB 1501 CTTGGGCGGAGACCTGCGGAGCTGCGGCTTGTGTGTATGAGAGCCAGGCGG 1560
QY 1561 GAGCGGCTCAGAGATGAGAGCAGCAGCAGCAGATGAATCACCACTTACACCA 1620
DB 1561 GAGCGGCTCAGAGATGAGAGCAGCAGCAGCAGATGAATCACCACTTACACCA 1620
QY 1621 CTTCTACAGATAGAGCCCTCCCAAGGCGCCACCTGCAATGAAGACCCGAC 1680
DB 1621 CTTCTACAGATAGAGCCCTCCCAAGGCGCCACCTGCAATGAAGACCCGAC 1680
QY 1681 CCGGACACCAAGGCAATTAATTCATTAATTAATTAATTAATTAATTAATTAAT 1740
DB 1681 CCGGACACCAAGGCAATTAATTCATTAATTAATTAATTAATTAATTAATTAAT 1740
QY 1741 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1800
DB 1741 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1800
QY 1801 CACAGAACTAACTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1859
DB 1801 CACAGAACTAACTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1859

RESULT 2
ID AAA63925 standard; DNA; 1438 BP.
XX
AC AAA63925;
XX
DT 04-DEC-2000 (first entry)
XX
DB DNA encoding protein related to Drosophila naked cuticle polypeptide.

XX Nkd gene, naked cuticle gene, segment-polarity gene; Wnt signalling;
 KW gene therapy; Nkd defect; cancer; ss.
 XX Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 8..1420
 FT /tag= a
 FT /product= "naked cuticle polypeptide"
 PN MO200049034-A1.
 PD 24-AUG-2000.
 PF 17-FEB-2000; 2000MO-US004188.
 PR 17-FEB-1999; 99US-0120646P.
 XX (STRD) UNITV LELAND STANFORD JUNIOR.
 PI Scott M, Zeng W, Wharton K;
 XX MPI: 2000-571967/53.
 DR P-PSDB; AAB08216.
 XX An isolated nucleic acid molecule useful for analyzing (genetic
 PT predisposition to) a disease state and for therapeutic purposes e.g.
 PT treatment of cancer comprises a sequence encoding a naked cuticle
 PT protein.
 XX
 XX Claim 3; Page 46-48; 58pp; English.
 CC The present sequence encodes a protein related to the Drosophila Nkd
 CC (naked cuticle) polypeptide. In Drosophila, Nkd is a segment-polarity
 CC gene whose expression is induced by Wnt signalling. The Nkd polypeptide
 CC acts to antagonize Wnt signalling. Nkd may link ion fluxes to the
 CC regulation of Wnt signal potency, duration or distribution. The Nkd
 CC polynucleotides can be used for identifying homologous or related
 CC proteins, to modulate the expression or function of Nkd polypeptides, and
 CC in studying associated physiological pathways. Nkd polynucleotides can
 CC also be used in gene therapy to treat disorders associated with Nkd
 CC defects. They may also be used for therapeutic purposes e.g. treatment of
 CC cancer.
 XX
 XX Sequence 1438 BP; 319 A; 478 C; 439 G; 202 T; 0 U; 0 Other;
 SQ
 Query Match 71.9%; Score 1336; DB 3; Length 1438;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 301 CAGCGAGATGAGAAGATGAGAGAGTGAAGCAACCTGCCAGGCTCCAGAAGCA 360
 QY 577 GCTGAAGTTGAAAGAGCTCCAGTGCAGCTGTCCATGAGAGAGACGCCGAGAGTGC 636
 Db 361 GCTGAAGTTGAAAGAGCTCCAGTGCAGCTGTCCATGAGAGAGAGAGAGAGTGC 420
 QY 637 GACCTTCACCTGTATGACTTTGACCAACAGCGCAAGGTCAACCCGAGAGCATCACAG 696
 Db 421 GACCTTCACCTGTATGACTTTGACCAACAGCGCAAGGTCAACCCGAGAGCATCACAG 480
 QY 697 CTGTGTCACACCATATGAGGTGTGTGACTCTCTGTCAACACTCCCAATCCAG 756
 Db 481 CTGTGTCACACCATATGAGGTGTGTGACTCTCTGTCAACACTCCCAATCCAG 540
 QY 757 CAAGATGCTGCGGGTAAAGCTCAACGTTGCCCCGATGAGCAAGCAAGAGAGCT 816
 Db 541 CAAGATGCTGCGGGTAAAGCTCAACGTTGCCCCGATGAGCAAGCAAGAGAGCT 600
 QY 817 CTTTGTCAATCAGAGCTGACCTGAGAGCGGCAAGGCCCGAGAGCAAGGCCACTGA 876
 Db 601 CTTTGTCAATCAGAGCTGACCTGAGAGCGGCAAGGCCCGAGAGCAAGGCCACTGA 660
 QY 877 GGAAGCTGCGGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 936
 Db 661 GGAAGCTGCGGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
 QY 937 CTTGAGAGAGTGTGCTGTACCAACATTCGTATGAGAGAGAGAGAGAGAGAGAG 996
 Db 721 CTTGAGAGAGTGTGCTGTACCAACATTCGTATGAGAGAGAGAGAGAGAGAGAG 780
 QY 997 CTAATTAGATCTGCGCGGGATAGAAAAGTACAGTCCCAATTTGGGCTGGCTCCCTTC 1056
 Db 781 CTAATTAGATCTGCGCGGGATAGAAAAGTACAGTCCCAATTTGGGCTGGCTCCCTTC 840
 QY 1057 CGTGGCCAGAAAGTCAAGAACTGCCCCCGCAGCTCCCAATCCCACTGATCTGCTCCA 1116
 Db 841 CGTGGCCAGAAAGTCAAGAACTGCCCCCGCAGCTCCCAATCCCACTGATCTGCTCCA 900
 QY 1117 TGAAGCGGAAGCATTCACATCCCAACCGAAAGCCCAAGCGGTGAGACCCGGCTCTT 1176
 Db 901 TGAAGCGGAAGCATTCACATCCCAACCGAAAGCCCAAGCGGTGAGACCCGGCTCTT 960
 QY 1177 CCACTTCTTTGAACCCCAATGCGCAAGGTCTAGAGTCTCACAACCGCTCCGGGGCAC 1236
 Db 961 CCACTTCTTTGAACCCCAATGCGCAAGGTCTAGAGTCTCACAACCGCTCCGGGGCAC 1020
 QY 1237 CCAAGACGGGAGCAAGCATTTGTGAGGTCCGCCAAGGCCAGAGGCAAGGTGAGGT 1296
 Db 1021 TCAAGACGGGAGCAAGCATTTGTGAGGTCCGCCAAGGCCAGAGGCAAGGTGAGGT 1080
 QY 1297 GGGCCAGTGGCCAGAGGGGCAAGAAACAAGCCCTTTGGGAACCGGCATCCCTGCGGT 1356
 Db 1081 GGGCCAGTGGCCAGAGGGGCAAGAAACAAGCCCTTTGGGAACCGGCATCCCTGCGGT 1140
 QY 1357 GTCCCCCTCCGCCCACTGGCTGCGCAAGCCCGGCGCTCTCCCTCTACAGCCCTCGG 1416
 Db 1141 GTCCCCCTCCGCCCACTGGCTGCGCAAGCCCGGCGCTCTCTCTCTACAGCCCTCGG 1200
 QY 1417 GCAAGAAGAGCAACAAGCAAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476
 Db 1201 GCAAGAAGAGCAACAAGCAAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 QY 1477 ACCACTGGCTCAGGTGGCCCTGTCTGTGGGGGGAGACCTGCGGAGCTGCCCCCTT 1536
 Db 1261 ACCACTGGCTCAGGTGGCCCTGTCTGTGGGGGGAGACCTGCGGAGCTGCCCCCTT 1320
 QY 1537 GGTGTGTATGAG 1596
 Db 1321 GGTGTGTATGAG 1380
 QY 1597 TGAACATCAACACCATTAACACCATTTCTACAGACATAGAGCCCTCCAGAGGCC 1654
 Db 1381 TGAACATCAACACCATTAACACCATTTCTACAGACATAGAGCCCTCCAGAGGCC 1438

RESULT 3
AAL03214
ID AAL03214 standard; DNA; 2379 BP.
XX
AC AAL03214;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 5902.
XX
KM Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN W0200155320-A2.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 23-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227182P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229309P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.

PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234597P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249267P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.

PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465570/50.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen is
XX PT used in preventing, treating or ameliorating a medical condition.
XX
XX Disclosure; SEQ ID NO 5902; 1297bp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
XX CC number of human reproductive system related antigens. These can be used
XX CC in the prevention and treatment of reproductive system disorders,
XX CC including cancer. The present sequence is a genomic sequence encoding a
XX CC protein of the invention
SQ Sequence 2379 BP; 548 A; 689 C; 539 G; 603 T; 0 U; 0 Other;
Query Match 42.1%; Score 783; DB 4; Length 2379;
Best Local Similarity 100.0%; Pred. No. 2.5e-284;
Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1046 GGGTCCCTCCCTCCGCGCCCAAGTCAAGACGCCCCCGGACCTTCAATCCCACTCGA 1105
DB 840 GGGTCCCTCCCTCCGCGCCCAAGTCAAGACGCCCCCGGACCTTCAATCCCACTCGA 899
QY 1106 TCTCGCTCCCATGAGCGGAGGACCATCCACACCGAAGCCCAAGGGGTGGAC 1165
DB 900 TCTCGCTCCCATGAGCGGAGGACCATCCACACCGAAGCCCAAGGGGTGGAC 959
QY 1166 CCGGCTCTCTTCACCTTCTTGAACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAAGG 1225
DB 960 CCGGCTCTCTTCACCTTCTTGAACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAAGG 1019
QY 1226 CTCCTGGGGGACCCAGAGCGGAGGACATTTGTGAGTCCCCCAAGGCCCAAGGCAAG 1285
DB 1020 CTCCTGGGGGACCCAGAGCGGAGGACATTTGTGAGTCCCCCAAGGCCCAAGGCAAG 1079
QY 1286 AGTGTGGGTGGGGACGTCGCGCAGAGGGGGCAAGAAAGGCCCTCTGGGAGCCCGGC 1345
DB 1080 AGTGTGGGTGGGGACGTCGCGCAGAGGGGGCAAGAAAGGCCCTCTGGGAGCCCGGC 1139
QY 1346 ATCCCTGCGGTGTCTCCCTTCGCGCACTTGTGCTGCGAGCCCGGCTCTCTCCCTCCCTA 1405
DB 1140 ATCCCTGCGGTGTCTCCCTTCGCGCACTTGTGCTGCGAGCCCGGCTCTCTCCCTCCCTA 1199
QY 1406 GCGCCCTCTCGGGACAAAGAACACACCGAGCCAGAGAGCCAGAGGGCTGCGG 1465
DB 1200 GCGCCCTCTCGGGACAAAGAACACACCGAGCCAGAGAGCCAGAGGGCTGCGG 1259
QY 1466 GGGCTGACGACACATGCGCTCAGGTGCGCTGTCTCTGGGGGGAGGACCTGCGGGAG 1525
DB 1260 GGGCTGACGACACATGCGCTCAGGTGCGCTGTCTCTGGGGGGAGGACCTGCGGGAG 1319
QY 1526 CTGCGCCGCTTGTGTGTATGAGAGCCAGGCGGAGCCGGTCCAGAGACATGAGCAC 1585
DB 1320 CTGCGCCGCTTGTGTGTATGAGAGCCAGGCGGAGCCGGTCCAGAGACATGAGCAC 1379

QY 1586 CACCACACCATGATGAATCAACCAACCATTAACCACTTCAACGAGACATAGAGCCCTCC 1645
DB 1380 CACCACACCATGATGAATCAACCAACCATTAACCACTTCAACGAGACATGAGCCCTCC 1439
QY 1646 CAGAGGCCCCACCTGCGCATATGAGAGCCCAACCCCGAGACACACAGGACATTAATT 1705
DB 1440 CAGAGGCCCCACCTGCGCATATGAGAGCCCAACCCCGAGACACACAGGACATTAATT 1499
QY 1706 CTATTAATTAATGTTATTAATGATGATTAATTAATTAATTAATTAATTAATTAAT 1765
DB 1500 CTATTAATTAATGTTATTAATGATGATTAATTAATTAATTAATTAATTAATTAAT 1559
QY 1766 ATTAGCTAGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1825
DB 1560 ATTAGCTAGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1619
QY 1826 GTT 1828
DB 1620 GTT 1622
RESULT 4
AAL03216
ID AAL03216 standard; DNA; 2379 BP.
XX
AC AAL03216;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 5904.
XX
KM Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-MAR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 16-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 23-AUG-2000; 2000US-0227182P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0234265P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236337P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.

PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250360P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
FA Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-46570/50.
XX
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.
XX
XX
XX Disclosure; SEQ ID NO 5904; 1297bp + Sequence Listing; English.
PS
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention
XX
SQ Sequence 2379 BP; 548 A; 689 C; 539 G; 603 T; 0 U; 0 Other;
Query Match 42.1%; Score 783; DB 4; Length 2379;
Best Local Similarity 100.0%; Pred. No. 2,5e-284;
Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1046 GGCCTCCCTCCGTCGCGCGGAGTCAGAACTGCCCCCGGCACTCCCATCCCACTCGA 1105
DB 840 GGCCTCCCTCCGTCGCGCGGAGTCAGAACTGCCCCCGGCACTCCCATCCCACTCGA 899
QY 1106 TCTCGCTCCCATGAGCGGAGCATCCATCCACACCGAAGGCCCAAGGCGTGGAC 1165
DB 900 TCTCGCTCCCATGAGCGGAGCATCCATCCACACCGAAGGCCCAAGGCGTGGAC 959
QY 1166 CCGGCTCTCTTCCACTTCTTGACACCCCAATGCGCAAGTCTTCAGAGCTTCAGCAACGG 1225
DB 960 CCGGCTCTCTTCCACTTCTTGACACCCCAATGCGCAAGTCTTCAGAGCTTCAGCAACGG 1019
QY 1226 CTCGGGGGACCCAGGACGGGAGCACTTGTGAGGTCCTCCCAAGGCCCAAGGCGCAAG 1285
DB 1020 CTCGGGGGACCCAGGACGGGAGCACTTGTGAGGTCCTCCCAAGGCCCAAGGCGCAAG 1079
QY 1286 AGTGTGGTGTGGGCACTGTGGCCAGAGGGGCAAGAAACAGGCCCTCTGGGACCCGCC 1345

```

Db      1080 AGGTGTGTGTGGGCGACGTGGCCAGAGGGGCAAGAAAGCCCTCTGGAGCCGCC 1139
Qy      1346 ATCCCTGGGTGTCCCTCCGCCACCTGTGGCTGCAGCCCGGCTCTCCCTCCCTA 1405
Db      1140 ATCCCTGGGTGTCCCTCCGCCACCTGTGGCTGCAGCCCGGCTCTCCCTCCCTA 1199
Qy      1406 GCGCCCTGGGGCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1465
Db      1200 GCGCCCTGGGGCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1259
Qy      1466 GCGCTGAGGCAACCACTGGCTCAGTGGCTGTCTGTGGGGCGGAGCACTGGCGGAG 1525
Db      1260 GCGCTGAGGCAACCACTGGCTCAGTGGCTGTCTGTGGGGCGGAGCACTGGCGGAG 1319
Qy      1526 CTGCGCGCTTGGTGTGTATGAGAGCCAGAGCGGGGAGCGGTCCAGAGACATGAGCAC 1585
Db      1320 CTGCGCGCTTGGTGTGTATGAGAGCCAGAGCGGGGAGCGGTCCAGAGACATGAGCAC 1379
Qy      1586 CACACACACATGAACATCAACCACTTACCACTTCTTACCAAGATAGAGCCCTCC 1645
Db      1380 CACACACACATGAACATCAACCACTTACCACTTCTTACCAAGATAGAGCCCTCC 1439
Qy      1646 CAGAGGCGCCACCTGCATATGAGAGCCGACCCCGACACCAAGGATATTATT 1705
Db      1440 CAGAGGCGCCACCTGCATATGAGAGCCGACCCCGACACCAAGGATATTATT 1499
Qy      1706 CTATTAATTAATTTATTTATGATGATTTATTTATTAATAATTAATTTATTTAT 1765
Db      1500 CTATTAATTAATTTATTTATGATGATTTATTTATTAATAATTAATTTATTTAT 1559
Qy      1766 ATTAGTAGCCTACATGTAGAGATCTATGAGAAACAGAGCTAACTTTTATTAT 1825
Db      1560 ATTAGTAGCCTACATGTAGAGATCTATGAGAAACAGAGCTAACTTTTATTAT 1619
Qy      1826 GTT 1828
Db      1620 GTT 1622

RESULT 5
AAL51525
ID      AAL51525 standard; DNA; 590 BP.
AC      AAL51525;
XX
XX
XX      24-APR-2003 (first entry)
DE      Human Nkd (hNkd) gene exon 11 (coding region).
XX
XX      Human; ds; gene therapy; Nkd; hNkd; Wnt signalling pathway; cancer;
KW      colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer;
KW      promoter.
XX
XX      Homo sapiens.
OS
XX
XX      WO200292832-A2.
PN
XX
XX      21-NOV-2002.
PD
XX
XX      27-NOV-2001; 2001WO-US044092.
PF
XX
XX      27-NOV-2000; 2000US-0252884P.
PR      16-MAY-2001; 2001US-0291109P.
PR      01-OCT-2001; 2001US-0325571P.
XX
XX      (CHIR ) CHIRON CORP.
PA
XX      Rohan M, Chan V, Yan D;
PI
XX      WPI; 2003-129303/12.
XX
XX      New human and non-human primate homologues of Nkd protein, and Nkd genes,

```

```

PT      useful for treating cancer involving aberrant Wnt signaling, e.g. colon
PT      cancer, head and neck cancer, ovarian cancer, or breast cancer.
XX
XX      Example 3; Fig 7; 99p; English.
XX
XX      The invention comprises the amino acid and coding sequence of the human
CC      Nkd protein (hNkd) - a regulator of Wnt signalling pathways. The hNkd DNA
CC      and protein sequences are useful for the treatment of cancer involving
CC      aberrant Wnt signaling (e.g. colon cancer, head and neck cancer, ovarian
CC      cancer and breast cancer). The present DNA sequence represents exon 11
XX      (coding region) from the human Nkd gene
XX
XX      Sequence 590 BP; 119 A; 234 C; 159 G; 78 T; 0 U; 0 Other;
SQ
Query Match      31.7%; Score 590; DB 10; Length 590;
Best Local Similarity 100.0%; Pred. No. 9,1e-212;
Matches 590; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1047 GCTCCCTTCCGGGCGCCAGAGTGAAGTGGCCCCCGGACCTCCATCCCATCTGGAT 1106
Db      1 GCTCCCTTCCGGGCGCCAGAGTGAAGTGGCCCCCGGACCTCCATCCCATCTGGAT 60
Qy      1107 CTGCTCCCATGAGCGGAAAGCCATCCATCCCAACCGAAAGCCCAAGGCGTGAGCC 1166
Db      61 CTGCTCCCATGAGCGGAAAGCCATCCATCCCAACCGAAAGCCCAAGGCGTGAGCC 120
Qy      1167 CGGCTCTCTTCACTTCTTTGACACCCCAATGCGCAAGTCTGAGCTTCAGCAAGCGC 1226
Db      121 CGGCTCTCTTCACTTCTTTGACACCCCAATGCGCAAGTCTGAGCTTCAGCAAGCGC 180
Qy      1227 TCGGGGCAACCCAGAGCGGAGCAAGCACTTTGTGAGTCTCCCAAGGCCCAAGGCAAGA 1286
Db      181 TCGGGGCAACCCAGAGCGGAGCAAGCACTTTGTGAGTCTCCCAAGGCCCAAGGCAAGA 240
Qy      1287 GTGTGGGTGTGGGCAAGTGGGCCAGAGAGGGCAAGAAACAAGCCCTCTGAGAACCCGCA 1346
Db      241 GTGTGGGTGTGGGCAAGTGGGCCAGAGAGGGCAAGAAACAAGCCCTCTGAGAACCCGCA 300
Qy      1347 TCCCTGGGTGTCCCTCCGCCACCTGGCTGCGAGCCCGGCTCTCTCCCTCTAG 1406
Db      301 TCCCTGGGTGTCCCTCCGCCACCTGGCTGCGAGCCCGGCTCTCTCCCTCTAG 360
Qy      1407 CCCCCCTCGGGCAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1466
Db      361 CCCCCCTCGGGCAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 420
Qy      1467 GCGTGCAGGCAACCACTGGCTCAGTGGCTGTGCTGTGGGGCGGAGCACTGGCGGAGC 1526
Db      421 GCGTGCAGGCAACCACTGGCTCAGTGGCTGTGCTGTGGGGCGGAGCACTGGCGGAGC 480
Qy      1527 TGCCCGCTTGTGTGTATGAGAGCCAGGCGGGGCAAGCGGTCCAGAGACATGAGCACC 1586
Db      481 TGCCCGCTTGTGTGTATGAGAGCCAGGCGGGGCAAGCGGTCCAGAGACATGAGCACC 540
Qy      1587 ACCACCAACATGAACATCAACCACTTACCACTTTCTACGAGACATAG 1636
Db      541 ACCACCAACATGAACATCAACCACTTACCACTTTCTACGAGACATAG 590

RESULT 6
AAS92082
ID      AAS92082 standard; cDNA; 1561 BP.
AC      AAS92082;
XX
XX      13-FEB-2002 (first entry)
DE
XX
XX      DNA encoding novel human diagnostic protein #27886.
XX
XX      Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW      food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX      Homo sapiens.
OS

```

[illegible]

Db	301	GGGAGAAAGACGAGGAGCCCGCTCAGGTTCAAGGTGACAGCCGGCTTGAGCAGTCTG	360
Qy	951	GCTGCTACCCACCATTCGGTAGATGAGAAACATCGAGAGGAAACCACTTAGATCTTG	101
Db	361	GCTGCTACCAACCATTCGCTAGTGTGAACATCGAGAGGAAACCACTTAGATCTTG	420
Qy	1011	CCGGATATGAAAACTACACGTCCTCCATTGTGGGCTTG	1046
Db	421	CCGGATATGAAAACTACACGTCCTCCATTGTGGGCTTG	456
RESULT 7			
ACH21474	ID	ACH21474 standard; cDNA, 427 BP.	
ACH21474;			
ACH21474;			
13-OCT-2003	(first entry)		
Human adult liver cDNA #1086.			
Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST; genome mapping; biodiversity; genetic disorder.			
Homo sapiens.			
US2003073623-A1.			
17-APR-2003.			
30-JUL-2001; 2001US-00918995.			
30-JUL-2001; 2001US-00918995.			
30-JUL-2001; 2001US-00918995.			
(DRWA/) DRMANAC R T.			
(LABA/) LABAT I.			
(STAC/) STRACHE-CRAIN B.			
(DICK/) DICKSON M C.			
(JONE/) JONES L W.			
Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;			
WPI; 2003-615964/58.			
New polynucleotide sequences obtained from various cDNA libraries, useful as hybridization probes, as oligomers for PCR, for chromosome and gene mapping, in the recombinant production of protein, or in generating antisense DNA or RNA.			
Claim 1; SEQ ID NO 8686; 44pp; English.			
The invention relates to an isolated polynucleotide comprising any one of 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was determined by the technique of SBH (sequencing by hybridisation). Also included is a purified polypeptide comprising a sequence corresponding to a reading frame of the novel polynucleotide. The nucleic acid sequences are useful in diagnostics as expressed sequence tags (EST) for identifying expressed genes or for physical mapping of the human genome, in forensics, in assessing biodiversity, or in identifying mutations responsible for genetic disorders and other traits. The nucleotide sequences are also useful as hybridisation probes, as oligomers for PCR, for chromosome and gene mapping, in the recombinant production of protein, or in generating antisense DNA or RNA. The purified polypeptide is useful for generating antibodies specific for it. The present sequence data is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPRO at segdata.uspto.gov/sequence.html?DocID=20030073623			
Sequence 427 BP; 84 A; 177 C; 112 G; 53 T; 0 U; 1 Other;			

Query Match 20.3%; Score 377; DB 9; Length 427;

Best Local Similarity 100.0%; Pred. No. 8,9e-132; Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1076 CTGCCCCCGGCACTTCCAAATCCCACTGATCTGCTCCCAATGAGCCGAAAGCCATCCAC 1135
Db 4 CTGCCCCCGGCACTTCCAAATCCCACTGATCTGCTCCCAATGAGCCGAAAGCCATCCAC 63

QY 1136 ATCCCAACCGAAAGCCGCAAGGCTGAGACCCGGCTCTTCCACTTCTTGAACACCCCA 1195
Db 64 ATCCCAACCGAAAGCCGCAAGGCTGAGACCCGGCTCTTCCACTTCTTGAACACCCCA 123

QY 1196 ATGCGCAAGGCTCAGAGCTCCAGCAACGAGCTCCGGGCAACCCAGAGCGGAGCAAC 1255
Db 124 ATGCGCAAGGCTCAGAGCTCCAGCAACGAGCTCCGGGCAACCCAGAGCGGAGCAAC 183

QY 1256 TTGTGAGGTCCCGCAAGGCGCCAGAGGCAAGTGTGAGTGTGGCCACTGTGGCAGAGG 1315
Db 184 TTGTGAGGTCCCGCAAGGCGCCAGAGGCAAGTGTGAGTGTGGCCACTGTGGCAGAGG 243

QY 1316 GCAGAAACCAAGCCCTCTGAGACCCGCCATCCCTGCGGTGTCTCCCTCCGCCACTG 1375
Db 244 GCAGAAACCAAGCCCTCTGAGACCCGCCATCCCTGCGGTGTCTCCCTCCGCCACTG 303

QY 1376 GCTGCGCAGCCCGGCTCTCTCCCTGAGCCCGCCCTCGGGCAAGAGCAAGCAAC 1435
Db 304 GCTGCGCAGCCCGGCTCTCTCCCTGAGCCCGCCCTCGGGCAAGAGCAAGCAAC 363

QY 1436 CGAGCCAGAGAGAGCA 1452
Db 364 CGAGCCAGAGAGAGCA 380

RESULT 8
AAS78752 ID AAS78752 standard; cDNA; 1743 BP.
XX AAS78752;
AC AAS78752;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #14556.
XX
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
XX P-PDB; ABG14565.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensic, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
PS Claim 1; SEQ ID NO 14556; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensic, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

SO Sequence 1743 BP; 367 A; 562 C; 511 G; 303 T; 0 U; 0 Other;

Query Match 11.7%; Score 218; DB 5; Length 1743;
Best Local Similarity 100.0%; Pred. No. 3.5e-72;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1047 GCTCCCTTCCTGCGGCCAGAAAGTCAAGTCCGCCCCCGCACTCCAAATCCACTGAT 1106
Db 332 GCTCCCTTCCTGCGGCCAGAAAGTCAAGTCCGCCCCCGCACTCCAAATCCACTGAT 391

QY 1107 CTGCGCTCCATGAGCCGGAAGCCATCCATCCCAACCGAAAGCCCAAGCGGTGAGC 1166
Db 392 CTGCGCTCCATGAGCCGGAAGCCATCCATCCCAACCGAAAGCCCAAGCGGTGAGC 451

QY 1167 CGGCTCTCTTCACTTCTTGAACCCCAATCGCAAGTCTCAGAGCTCAGCAACGGC 1226
Db 452 CGGCTCTCTTCACTTCTTGAACCCCAATCGCAAGTCTCAGAGCTCAGCAACGGC 511

QY 1227 TCCGGGGCAACCGAGCGGAGCAAGCACTTTGTGAGG 1264
Db 512 TCCGGGGCAACCGAGCGGAGCAAGCACTTTGTGAGG 549

RESULT 9
AAS92081 ID AAS92081 standard; cDNA; 1743 BP.
XX AAS92081;
AC AAS92081;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #27885.
XX
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
XX P-PDB; ABG27894.
XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

PS Claim 1; SEQ ID NO 27885; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantifying a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1743 BP; 367 A; 562 C; 511 G; 303 T; 0 U; 0 Other;

XX Query Match 11.7%; Score 218; DB 5; Length 1743;

XX Best Local Similarity 100.0%; Pred. No. 3.5e-72; Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1047 GCTCCCTTCCTCGGCCAGAGTCAAGTGTGCCCCCGGACCTCCCAATCCCATCTGAT 1106

Db 332 GCTCCCTTCCTCGGCCAGAGTCAAGTGTGCCCCCGGACCTCCCAATCCCATCTGAT 391

QY 1107 CTGCTCCCATGAGCCGGAAGCCATCCATCCCAACCGAAGCCCAAGGCGTGAAC 1166

Db 392 CTGCTCCCATGAGCCGGAAGCCATCCATCCCAACCGAAGCCCAAGGCGTGAAC 451

QY 1167 CGGCTCTCTTCCACTTCTTGACACCCCATCGGCAAGGTCTCAGAGCTCCAGCAAGGC 1226

Db 452 CGGCTCTCTTCCACTTCTTGACACCCCATCGGCAAGGTCTCAGAGCTCCAGCAAGGC 511

QY 1227 TCCGGGGACACCGAGAGCGGAGCACTTTGTAGG 1264

Db 512 TCCGGGGACACCGAGAGCGGAGCACTTTGTAGG 549

RESULT 10

ADCC1966

ID ADCC1966 standard; cDNA; 1743 BP.

XX ADCC1966;

XX 18-DEC-2003 (first entry)

XX Human novel cDNA contig sequence, SEQ ID NO:2048.

XX Human; diagnostic; drug screening; forensics; gene mapping;

XX biodiversity assessment; Parkinson's disease; Alzheimer's disease;

XX neurodegenerative diseases; anaemia; platelet disorders; wound; burns;

XX ulcers; osteoporosis; autoimmune disease; cancer;

XX molecular weight marker; food supplement; antiparkinsonian; nootropic;

XX neuroprotective; anti-neumatic; anticoagulant; thrombolytic; vlnnerary;

XX anticler; osteopathic; immunosuppressive; antiinflammatory; cytostatic;

XX gene therapy; se.

XX Homo sapiens.

XX OS

XX

PN WO2003029271-A2.

XX 10-APR-2003.

XX 24-SEP-2002; 2002WO-US030474.

XX 24-SEP-2001; 2001US-0324631P.

XX (HYSEQ) HYSEQ INC.

XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;

XX Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;

XX Haley-Vicente D, Drmanac RT;

XX WPI; 2003-371981/35.

XX P-PSDB; ADCC2733.

XX Example 2; SEQ ID NO 2048; 1185bp; English.

XX The invention relates to 971 novel human cDNA sequences (ADCC29919-
CC ADCC0889) and the polypeptides they encode (ADCC0890-ADCC1860). The
CC invention also relates to nucleic acid sequences over 99% identical with
CC the novel human cDNAs. The invention additionally encompasses expression
CC vectors and host cells comprising a nucleic acid of the invention; the
CC recombinant production of a polypeptide of the invention; an antibody
CC against a polypeptide of the invention; a method of detecting
CC polynucleotides or polypeptides of the invention; and methods of
CC identifying a compound which binds to a polypeptide of the invention. The
CC invention further discloses methods of preventing, treating or
CC ameliorating a medical condition; kits comprising polynucleotide probes
CC and/or monoclonal antibodies for carrying out the methods of the
CC invention; methods for the identification of compounds that modulate the
CC expression or activity of the polynucleotide and/or polypeptide; and 767
CC contig sequences corresponding to the cDNA sequences of the invention
CC (ADCC1861-ADCC2627) and the polypeptides encoded by the contigs (ADCC2628
CC -ADCC3394). The nucleic acids and polypeptides of the invention are
CC useful in diagnostics, drug screening, forensics, gene mapping, in the
CC identification of mutations responsible for genetic disorders or other
CC traits, for assessing biodiversity, and in producing many other types of
CC data and products dependent on DNA and amino acid sequences. They are
CC also used for treating diseases such as Parkinson's disease, Alzheimer's
CC disease and other neurodegenerative diseases, anaemia, platelet
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primers, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a human contig
CC sequence used in an example of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1743 BP; 367 A; 562 C; 511 G; 303 T; 0 U; 0 Other;

XX Query Match 11.7%; Score 218; DB 10; Length 1743;

XX Best Local Similarity 100.0%; Pred. No. 3.5e-72; Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1047 GCTCCCTTCCTCGGCCAGAGTCAAGTGTGCCCCCGGACCTCCCAATCCCATCTGAT 1106

Db 332 GCTCCCTTCCTCGGCCAGAGTCAAGTGTGCCCCCGGACCTCCCAATCCCATCTGAT 391

QY 1107 CTGCTCCCATGAGCCGGAAGCCATCCATCCCAACCGAAGCCCAAGGCGTGAAC 1166

Db 392 CTGCTCCCATGAGCCGGAAGCCATCCATCCCAACCGAAGCCCAAGGCGTGAAC 451

QY 1167 CGGCTCTCTTCCACTTCTTGACACCCCATCGGCAAGGTCTCAGAGCTCCAGCAAGGC 1226

XX Human genome derived single exon probe #7990.
 DE Human; probe, ss; gene expression; single exon probe; microarray;
 KW alternative splicing event; genomic alteration.
 XX Homo sapiens.
 OS
 PN US2003194704-A1.
 PD 16-OCT-2003.
 XX
 PF 03-APR-2002; 2002US-00029386.
 XX
 PR 03-APR-2002; 2002US-00029386.
 XX
 PA (PENN/) PENN S G.
 PA (RANK/) RANK D R.
 PA (HANZ/) HANZEL D K.
 XX
 PI Penn SG, Rank DR, Hanzel DK,
 DR WPI; 2004-119264/12.
 XX
 PT New human genome-derived single exon nucleic acid probes useful for human
 PT gene expression analysis, for identifying or characterizing alternative
 PT splicing events, for assessing genomic alterations or as tools for
 PT surveying tissues.
 PS
 PS Claim 15; SEQ ID NO 7990; 80pp; English.
 XX
 XX The invention relates to a nucleic acid probe for measuring human gene
 CC expression, comprising any of the 27,400 fully defined nucleotide
 CC sequences in the specification, or their complements or fragments, and
 CC encoding at least 8 amino acids of any of the 688 amino acid sequences
 CC fully defined in the specification. The probe is a single exon probe that
 CC hybridizes under high stringency conditions to a nucleic acid molecule
 CC expressed in human cells or tissues. Also included are a spatially-
 CC addressable set of single exon nucleic acid probes for measuring human
 CC gene expression (comprising a plurality of single exon nucleic acid
 CC probes cited above, where each of the plurality of probes is separately
 CC and addressably isolatable or amplifiable from the plurality), a single
 CC exon microarray for measuring human gene expression, a method of
 CC measuring human gene expression, a vector comprising the single exon
 CC probe cited above, an ORF-encoded peptide comprising at least 8
 CC contiguous amino acids of any of the above-mentioned amino acid
 CC sequences (optionally with conservative amino acid substitutions), an
 CC isolated antibody that binds specifically to a peptide cited above,
 CC methods of selling and/or licensing single exon probes or microarrays to
 CC a customer desiring to measure gene expression, a method of providing
 CC human gene expression data by subscription, and a computer-readable
 CC storage medium which contains a database having a plurality of records
 CC (each record including data on the expression of a single exon probe
 CC cited above. The probe, methods and apparatus are useful in gene
 CC expression analysis. The probes may be used as tools for surveying
 CC tissues to detect the presence of expressed messages that contain their
 CC specific exon, or in constructing genome-derived single exon microarrays.
 CC In addition, the probes are used in identifying and characterizing
 CC alternative splicing events, in detecting and characterizing gross
 CC alterations in the genomic locus that includes their exon, in assessing
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,
 CC or in expressing the ORF-encoded peptide. The present sequence is a human
 CC single exon probe of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?DocID=20030194704
 XX
 SQ Sequence 598 BP; 84 A; 232 C; 162 G; 120 T; 0 U; 0 Other;
 Query Match 7.4%; Score 137; DB 12; Length 598;
 Best Local Similarity 100.0%; Pred. No. 1.1e-41;
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 AGGTGACAGCTTCGCGCTGAGCCGCTGCGGCTCGGAGAGGATCGAGATCGATCGG 339
 |||||
 DB 228 AGGTGACAGCTTCGCGCTGAGCCGCTGCGGCTCGGAGAGGATCGAGATCGG 287
 QY 340 GAGACAGCGCTGCCCGGCGGTGTCTTCGGGACCCCGACAGCTGCGTTGCGGCGGACCAT 399
 |||||
 DB 288 GAGACAGCGCTGCCCGGCGGTGTCTTCGGGACCCCGACAGCTGCGTTGCGGCGGACCAT 347
 QY 400 AGGCCGAGACGACCCGGG 416
 |||||
 DB 348 AGGCCGAGACGACCCGGG 364
 RESULT 14
 ID ABQ73820 standard; cDNA; 1950 BP.
 XX
 AC ABQ73820;
 XX
 DT 07-OCT-2002 (first entry)
 DE Human colon specific nucleic acid (CSNA) SEQ ID NO:126.
 XX
 KW Human; colon specific nucleic acid; colon specific polypeptide; CSP;
 KW CSNA; colon specific gene; CSG; colon cancer; gene therapy; vaccine;
 KW cytostatic; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200248370-A2.
 XX
 PD 20-JUN-2002.
 XX
 PF 30-OCT-2001; 2001WO-US051341.
 XX
 PR 31-OCT-2000; 2000US-0244717P.
 XX
 PA (DIAD-) DIADEXUS INC.
 XX
 PI Sun Y, Reipon H, Ghosh MG, Liu C;
 XX
 DR WPI; 2002-583520/62.
 XX
 PT Colon specific polypeptides and polynucleotides useful for detecting,
 PT diagnosing, monitoring, treating, staging and predicting cancers in
 PT humans having cancer and non-cancerous colon disease.
 PS
 PS Claim 1; Page 199; 243pp; English.
 XX
 CC ABQ73695 to ABQ73841 represent human colon specific nucleic acid (CSNA)
 CC sequences, and ABP51826 to ABP51928 represent human colon specific
 CC polypeptide (CSP) sequences from the present invention. CSNA and CSP
 CC sequences have cytostatic activity, and can be used in gene therapy,
 CC antisense therapy and in vaccines. CSNA and CSP sequences can be used for
 CC diagnosing and monitoring the presence and metastases of colon cancer in
 CC a patient, by determining an amount of CSP or CSNA in a sample of a
 CC patient, and comparing it to the amount of colon specific marker in a
 CC normal control, where a difference in the amount of the nucleic acid or
 CC the polypeptide in the sample compared to that of normal control is
 CC associated with presence of colon cancer. CSP and CSNA sequences can be
 CC used for producing engineered colon tissue for treatment and research.
 CC CSNA sequences are useful for producing transgenic animals and cells and
 CC also in gene therapy
 XX
 SQ Sequence 1950 BP; 459 A; 494 C; 530 G; 467 T; 0 U; 0 Other;
 Query Match 7.4%; Score 137; DB 6; Length 1950;
 Best Local Similarity 100.0%; Pred. No. 8.8e-42;
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 281 GGTGACAGCTTCGCCGCTGAGCCGCTGCGGCTCGGAGAGGATCGAGATCGATCGG 340
 |||||
 DB 838 GGTGACAGCTTCGCCGCTGAGCCGCTGCGGCTCGGAGAGGATCGAGATCGATCGG 897

Search completed: December 29, 2004, 21:21:14
Job time : 1271 secs

QY 341 AGACAGCGCTGCGGCGGCTGCTCGGAGACCCGAGACGCTGCGGCTTGGCGGCAACCATTA 400
DB 898 AGACAGCGCTGCGGCGGCTGCTCGGAGACCCGAGACGCTGCGGCTTGGCGGCAACCATTA 957
QY 401 GCGCGAAGCACCAGGGA 417
DB 958 GCGCGAAGCACCAGGGA 974

RESULT 15

AL51518
ID AAL51518 standard; DNA; 134 BP.

AC AAL51518;

DT 24-APR-2003 (first entry)

DE Human Nkd (hNkd) gene exon 4 (coding region).

KW Human; ds; gene therapy; Nkd; hNkd; Wnt signalling pathway; cancer;

KW colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer;

OS Homo sapiens.

PN WO200292832-A2.

PD 21-NOV-2002.

PF 27-NOV-2001; 2001WO-US044092.

PR 27-NOV-2000; 2000US-0252884P.

PR 16-MAY-2001; 2001US-0291109P.

PR 01-OCT-2001; 2001US-0325571P.

PA (CHIR) CHIRON CORP.

PI Rohan M, Chan V, Yan D;

DR WPI; 2003-129303/12.

PT New human and non-human primate homologues of Nkd protein, and Nkd genes,

PT useful for treating cancer involving aberrant Wnt signaling, e.g. colon

PT cancer, head and neck cancer; ovarian cancer, or breast cancer.

PS Example 3; Fig 7; 99p; English.

CC The invention comprises the amino acid and coding sequence of the human

CC Nkd protein (hNkd) - a regulator of Wnt signalling pathways. The hNkd DNA

CC and protein sequences are useful for the treatment of cancer involving

CC aberrant Wnt signalling (e.g. colon cancer, head and neck cancer, ovarian

CC cancer and breast cancer). The present DNA sequence represents exon 4

CC (coding region) from the human Nkd gene

CC

QY Sequence 134 BP; 21 A; 40 C; 55 G; 18 T; 0 U; 0 Other;

DB

QY Query Match 7.2%; Score 134; DB 10; Length 134;

DB Best Local Similarity 100.0%; Pred. No. 1.8e-40; Mismatches 0; Indels 0; Gaps 0;

DB Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 282 GTGACAGCTTGGCGGCTGCTCGGAGACCCGAGACGCTGCGGCTTGGCGGCAACCATTA 341

DB 1 GTGACAGCTTGGCGGCTGCTCGGAGACCCGAGACGCTGCGGCTTGGCGGCAACCATTA 60

QY 342 GACAGCGCTGCGGCGGCTGCTCGGAGACCCGAGACGCTGCGGCTTGGCGGCAACCATTA 401

DB 61 GACAGCGCTGCGGCGGCTGCTCGGAGACCCGAGACGCTGCGGCTTGGCGGCAACCATTA 120

QY 402 GCCGAAAGCACCAGG 415

DB 121 GCCGAAAGCACCAGG 134

THIS PAGE IS BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: December 29, 2004, 20:56:19 ; Search time 212 Seconds
(Without alignments)
6232.814 Million cell updates/sec

Title: US-09-993-966-5

Perfect score: 1859
Sequence: 1 gacatcgccctctatacgcg.....aaaaaaaaaagcgccgc 1859

Scoring table: ~~OVERGROWN~~
Gapop 60.0 , Gapext 60.0

Searched: 824507 seqs, 355394441 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/5A COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B COMB.seq: *
5: /cgn2_6/prodata/1/ina/PCUS COMB.seq: *
6: /cgn2_6/prodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Query Match Length	DB ID	Description
1	1336	71.9	1438 4	US-09-506-066E-5 Sequence 5, Appl
2	53	2.9	3768 4	US-10-101-464A-949 Sequence 949, App
3	50	2.7	1731 4	US-09-506-066E-3 Sequence 3, Appl
4	45	2.4	45 4	US-09-686-583B-65 Sequence 65, Appl
5	45	2.4	45 4	US-09-672-725C-18 Sequence 18, Appl
6	31	1.7	47 2	US-08-778-494B-114 Sequence 114, App
7	31	1.7	50 1	US-08-381-572-20 Sequence 20, Appl
8	31	1.7	50 1	US-08-592-820-20 Sequence 20, Appl
9	31	1.7	50 4	US-09-129-603-3 Sequence 3, Appl
10	31	1.7	51 2	US-08-582-562A-8 Sequence 8, Appl
11	31	1.7	51 2	US-08-778-494B-8 Sequence 8, Appl
12	31	1.7	51 2	US-08-859-998-1373 Sequence 1373, Ap
13	31	1.7	51 3	US-09-225-928-1373 Sequence 1373, Ap
14	31	1.7	51 4	US-09-225-928-1373 Sequence 1373, Ap
15	31	1.7	52 4	US-08-618-100B-9 Sequence 9, Appl
16	31	1.7	52 4	US-09-807-784B-8 Sequence 8, Appl
17	31	1.7	55 2	US-08-522-562B-18 Sequence 18, Appl
18	31	1.7	55 2	US-08-582-562A-16 Sequence 16, Appl
19	31	1.7	55 3	US-08-778-494B-16 Sequence 16, Appl
20	31	1.7	55 3	US-09-294-923-18 Sequence 18, Appl
21	31	1.7	60 1	US-08-241-465B-11 Sequence 11, Appl
22	31	1.7	165 2	US-08-783-395-3 Sequence 3, Appl
23	31	1.7	165 2	US-08-924-83B-8 Sequence 8, Appl
24	31	1.7	471 5	PCT-US95-13658-1 Sequence 1, Appl
25	31	1.7	482 4	US-09-843-472-1 Patent No. 5185441
26	31	1.7	855 6	5185441-40 Patent No. 5223394
27	31	1.7	855 6	5223394-3 Patent No. 5223394

28	31	1.7	863 1	US-07-940-861-11 Sequence 11, Appl
29	31	1.7	863 2	US-08-459-512-11 Sequence 11, Appl
30	31	1.7	863 2	US-08-459-557-11 Sequence 11, Appl
31	31	1.7	863 2	US-08-460-132-11 Sequence 11, Appl
32	31	1.7	863 5	PCT-US92-02050-11 Patent No. 5185441
33	31	1.7	863 6	5185441-35 Patent No. 5185441
34	31	1.7	863 6	5223394-5 Patent No. 5223394
35	31	1.7	1003 2	US-08-887-997B-1 Sequence 1, Appl
36	31	1.7	1024 4	US-09-328-475C-9 Sequence 9, Appl
37	31	1.7	1024 4	US-09-328-475C-20 Sequence 20, Appl
38	31	1.7	1024 4	US-09-328-475C-37 Sequence 37, Appl
39	31	1.7	1338 2	US-08-484-933B-7 Sequence 7, Appl
40	31	1.7	1338 2	US-08-484-158B-7 Sequence 7, Appl
41	31	1.7	1338 2	US-08-484-596A-7 Sequence 7, Appl
42	31	1.7	1338 2	US-08-480-150A-7 Sequence 7, Appl
43	31	1.7	1338 3	US-08-458-731-7 Sequence 7, Appl
44	31	1.7	1338 3	US-08-149-223A-7 Sequence 7, Appl
45	31	1.7	1502 2	US-08-651-940-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-506-066E-5
Sequence 5, Application US/09506066E
Patent No. 6630323
GENERAL INFORMATION:
APPLICANT: Scott, Matthew
APPLICANT: Wharton, Keith
APPLICANT: Zeng, Wenlin
TITLE OR INVENTION: Naked Cuticle Genes and their Uses
FILE REFERENCE: STAN-121
CURRENT APPLICATION NUMBER: US/09/506,066E
CURRENT FILING DATE: 2000-02-17
PRIORITY APPLICATION NUMBER: 60/120,646
PRIORITY FILING DATE: 1999-02-17
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 1438
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (8)...(1418)
OTHER INFORMATION: Nkdl coding sequence
US-09-506-066E-5

Query Match 71.9%; Score 1336; DB 4; Length 1438;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	217	CCCCAGATGGGGAACCTTCACTCAAGCCGCGCGCTGTGAAGGAGGAGGCC	276
DB	1	CCCCAGATGGGGAACCTTCACTCAAGCCGCGCGCTGTGAAGGAGGAGGCC	60
QY	277	GGAAGGTGACAGCTTGCCCTGAGCCCTGCTGGATCGGAAGGATCGAGATGAT	336
DB	61	GGAAGGTGACAGCTTGCCCTGAGCCCTGCTGGATCGGAAGGATCGAGATGAT	120
QY	337	CGGAGACAGCGCTGCGCGCGGTCTTCGGAACCCGACAGCTCGGTCGGGAC	396
DB	121	CGGAGACAGCGCTGCGCGCGGTCTTCGGAACCCGACAGCTCGGTCGGGAC	180
QY	397	CATAGCCCAAGACCCGGAAGCTCGTGGCCGATGAGACACCGCTCAGCGAGA	456
DB	181	CATAGCCCAAGACCCGGAAGCTCGTGGCCGATGAGACACCGCTCAGCGAGA	240
QY	457	AGAGGAGACGACTTTCGCTGGAATGGCCCTGCTCGTGAAGACTGACGGGCTGG	516
DB	241	AGAGGAGACGACTTTCGCTGGAATGGCCCTGCTCGTGAAGACTGACGGGCTGG	300

51	CAGCGGAGATGAGAAGAAATGAGAGAGTGAAGGAACCTCGGCCAGGCTCAAGAAACA	576
57	CAGCGGAGATGAGAAGAAATGAGAGAGTGAAGGAACCTCGGCCAGGCTCAAGAAACA	576
Db	301 CAGCGGAGATGAGAAGAAATGAGAGAGTGAAGGAACCTCGGCCAGGCTCAAGAAACA	360
QY	577 GCTGAAGTTTGAAAGAGCTCCAGTGCAGCTGTCTCATGAGGAGACAGCCGGCAGAGTG	636
Db	361 GCTGAAGTTTGAAAGAGCTCCAGTGCAGCTGTCTCATGAGGAGACAGCCGGCAGAGTG	420
QY	637 GACCTTCAACCTGTATGACCTTTGACAAACACGACAGGTCAACCCGAGAGACATCACAG	696
Db	421 GACCTTCAACCTGTATGACCTTTGACAAACACGACAGGTCAACCCGAGAGACATCACAG	480
QY	697 CTTGTGTGACACCACTTATGAGGTGTGATCTCCTGTCAACCACTCCCAATCCAG	756
Db	481 CTTGTGTGACACCACTTATGAGGTGTGATCTCCTGTCAACCACTCCCAATCCAG	540
QY	757 CAAGATGTCTGGGGTAAAGCTCACCGTGGCCCCGATGGCAGCCAGACAAAGAGAGCT	816
Db	541 CAAGATGTCTGGGGTAAAGCTCACCGTGGCCCCGATGGCAGCCAGACAAAGAGAGCT	600
QY	817 CTTGTCAATCAGGCTGACCTTGACAGCGCAAGGCCCGAGCAAGACCAAGCCACTGA	876
Db	601 CTTGTCAATCAGGCTGACCTTGACAGCGCAAGGCCCGAGCAAGACCAAGCCACTGA	660
QY	877 GGACCTGGGAGCTGGGAGGAAGAACAGCCGCCCCGTCAAGTTTCCAGGGTGCAGCCG	936
Db	661 GGACCTGGGAGCTGGGAGGAAGAACAGCCGCCCCGTCAAGTTTCCAGGGTGCAGCCG	720
QY	937 CCTGAGACAGTCTGGGCTCTACCAACATTTGCTAGATAGAAACATCGAGAGAAACCA	996
Db	721 CCTGAGACAGTCTGGGCTCTACCAACATTTGCTAGATAGAAACATCGAGAGAAACCA	780
QY	997 CTACTTAGATCTCGCGGGATGAAACCTAACACGTCCCAATTTGGGCTGGCTCCCTTC	1056
Db	781 CTACTTAGATCTCGCGGGATGAAACCTAACACGTCCCAATTTGGGCTGGCTCCCTTC	840
QY	1057 CGTGGCCCAAGTTCAGAACTGCCCCCGGACCTCCAAATCCCACTGATTTGGTCCCA	1116
Db	841 CGTGGCCCAAGTTCAGAACTGCCCCCGGACCTCCAAATCCCACTGATTTGGTCCCA	900
QY	1117 TGAGCCGGAGACATCCATCCCAACGAAAGCCCCCAAGGCGTGGACCCGGCCTCCTT	1176
Db	901 TGAGCCGGAGACATCCCAATCCCAACGAAAGCCCCCAAGGCGTGGACCCGGCCTCCTT	960
QY	1177 CCACTTCTTTGACACCCCAATCGCCMAAGTCTCAGAGCTCCAGCAACGGCTCCGGAGCAC	1236
Db	961 CCACTTCTTTGACACCCCAATCGCCMAAGTCTCAGAGCTCCAGCAACGGCTCCGGAGCAC	1020
QY	1237 CCAGGACGGGAGCAAGCACTTTGTATGAGTCCCCCAAGGCCCAAGGCAAGATGTGGGTGT	1296
Db	1021 TCAGGACGGGAGCAAGCACTTTGTATGAGTCCCCCAAGGCCCAAGGCAAGATGTGGGTGT	1080
QY	1297 GGGCCAGGTGGACAGAGGGGGCAAGAAACAGAGCCCCCTGTGGAGCCCGGCATCCCTGCGGT	1356
Db	1081 GGGCCAGGTGGACAGAGGGGGCAAGAAACAGAGCCCCCTGTGGAGCCCGGCATCCCTGCGGT	1140
QY	1357 GTCCCCCTCGGCCCACTTGCTGACAGCCCGGCTCTCTCCCTTCCATAGCCCCCTTCGG	1416
Db	1141 GTCCCCCTCGGCCCACTTGCTGACAGCCCGGCTCTCTCCCTTCCATAGCCCCCTTCGG	1200
QY	1417 GCACAAAGAGCAACAGCAACCGAGCCAGAGAGCCAGAGGCTGCGGGGCTTGAGGC	1476
Db	1201 GCACAAAGAGCAACAGCAACCGAGCCAGAGAGCCAGAGGCTGCGGGGCTTGAGGC	1260
QY	1477 ACGACTGGCCATCAGGTGGCCCTGTCTGTGGGCGGGAGACCTTGCGGGAGCTGCCGCTT	1536
Db	1261 ACGACTGGCCATCAGGTGGCCCTGTCTGTGGGCGGGAGACCTTGCGGGAGCTGCCGCTT	1320
QY	1537 GGTGGTGTATAGAGCCCAAGCCCGGCAAGCCGGTTCAGAGACATGAGACCAACCA	1596
Db	1331 GGTGGTGTATAGAGCCCAAGCCCGGCAAGCCGGTTCAGAGACATGAGACCAACCA	1380
QY	1597 TGAACATCACCACTTACACACTTCTAACAGCATATAGACCCCTTCCAGGGGCC	1654

```

Db      1381  TGAATATCACCACCTTACCACTTCTTACACGACATATGAGCCCTCCCGAGGCC 1438

RESULT 2
US-10-101-464A-949
; Sequence 949, Application US/10101464A
; Patent No. 6768041
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; APPLICANT: Higgins, Colleen M.
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; FILE REFERENCE: 11000.1020C2
; CURRENT APPLICATION NUMBER: US/10/101,464A
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 949
; LENGTH: 3768
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-10-101-464A-949

Query Match      2.9%; Score 53; DB 4; Length 3768;
Best Local Similarity 100.0%; Pred.No. 7,1e-12;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      13  CTAATACGACTCTACTATAGGGGACAGCTGTGTAAACACCGCAAGTACGGGGG 65
         |||||
Db      1  CTAATACGACTCTACTATAGGGGACAGCTGTGTAAACACCGCAAGTACGGGGG 53

RESULT 3
US-09-506-066E-3
; Sequence 3, Application US/09506066E
; Patent No. 6630323
; GENERAL INFORMATION:
; APPLICANT: Scott, Matthew
; APPLICANT: Wharton, Keith
; APPLICANT: Zeng, Wenlin
; TITLE OF INVENTION: Naked Cuticle Genes and their Uses
; FILE REFERENCE: STAN-121
; CURRENT APPLICATION NUMBER: US/09/506,066E
; CURRENT FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/120,646
; PRIOR FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (140)..(1553)
; OTHER INFORMATION: Nkdl coding sequence
US-09-506-066E-3

Query Match      2.7%; Score 50; DB 4; Length 1731;
Best Local Similarity 100.0%; Pred.No. 1,2e-10;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      242  AAGCCGCGCCGCTGTGCAAGCGCAGGAGAGCCCGGAAAGGTACAGCTT 251

```

Db 158 AACCCGCGCCGCTGTGCAAGCGGAGAGCCCGGAGGTGACACTT 207

RESULT 4

US-09-686-583B-65
; Sequence 65, Application US/09666583B
; Patent No. 6576750
; GENERAL INFORMATION:
; APPLICANT: Heska Corporation
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Mismewski, Nancy
; TITLE OF INVENTION: FLU PERITOPHIN NUCLEIC ACID MOLECULES, PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-6-C2
; CURRENT APPLICATION NUMBER: US/09/686,583B
; PRIOR FILING DATE: 2000-10-11
; PRIOR APPLICATION NUMBER: 09/543,668
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128,704
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 65
; LENGTH: 45
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Primer
US-09-686-583B-65

Query Match 2.4%; Score 45; DB 4; Length 45;

Best Local Similarity 100.0%; Pred. No. 1.6e-08;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 CTATACGACTCACTATAGGGCAAGCAGTGTAAACAAGCAGAGT 57
Db 1 CTATACGACTCACTATAGGGCAAGCAGTGTAAACAAGCAGAGT 45

RESULT 5

US-09-672-725C-18
; Sequence 18, Application US/09672725C
; Patent No. 6753177
; GENERAL INFORMATION:
; APPLICANT: Stocker, Penny J.
; APPLICANT: Steimel-Crespi, Dorothy T.
; APPLICANT: Crespi, Charles L.
; APPLICANT: Rief, Timothy C.
; APPLICANT: Patten, Christopher J.
; TITLE OF INVENTION: P-GLYCOPROTEINS AND USES THEREOF
; FILE REFERENCE: G0307/7017
; CURRENT APPLICATION NUMBER: US/09/672,725C
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/156,510
; PRIOR FILING DATE: 1999-09-28
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 45
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic oligonucleotide
US-09-672-725C-18

Query Match 2.4%; Score 45; DB 4; Length 45;

Best Local Similarity 100.0%; Pred. No. 1.6e-08;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 CTATACGACTCACTATAGGGCAAGCAGTGTAAACAAGCAGAGT 57
Db 1 CTATACGACTCACTATAGGGCAAGCAGTGTAAACAAGCAGAGT 45

US-08-778-494B-114/c

; Sequence 114, Application US/08778494B
; Patent No. 5962272

GENERAL INFORMATION:

APPLICANT: Chenchik, Alex
APPLICANT: Zhu, York
APPLICANT: Diachenko, Luda
APPLICANT: Siebert, Paul
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR FULL-LENGTH CDNA
TITLE OF INVENTION: CLONING
NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/778,494B
FILING DATE: 03-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/582,562
FILING DATE: 03-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Doran R.
REGISTRATION NUMBER: 38,261
REFERENCE/DOCKET NUMBER: CL-7C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (352) 375-8100
TELEFAX: (352) 372-5800
INFORMATION FOR SEQ ID NO: 114:

SEQUENCE CHARACTERISTICS:
LENGTH: 47 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (synthetic)
US-08-778-494B-114

Query Match 1.7%; Score 31; DB 2; Length 47;
Best Local Similarity 100.0%; Pred. No. 0.0033;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1829 AAAAAAAAAAAAAAAAAAAGCGCGCCG 1859
Db 38 AAAAAAAAAAAAAAAAAAAGCGCGCCG 8

RESULT 7

US-08-381-572-20/c
; Sequence 20, Application US/08381572
; Patent No. 5565340

GENERAL INFORMATION:

APPLICANT: Chenchik, Alex
APPLICANT: Diachenko, Luda
APPLICANT: Siebert, Paul
APPLICANT: Lukianov, Sergey
APPLICANT: Gurskaya, Nadia
APPLICANT: Tarabukin, Victor

APPLICANT: Lukianov, Konstantin
TITLE OF INVENTION: METHOD FOR SUPPRESSING DNA FRAGMENT
TITLE OF INVENTION: AMPLIFICATION DURING PCR
NUMBER OF SEQUENCES: 26

```

CORRESPONDENCE ADDRESS:
ADDRESS: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/381,572
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Doran R.
REGISTRATION NUMBER: 38,361
REFERENCE/DOCKET NUMBER: CL-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (synthetic)
US-08-381-572-20

Query Match      1.7%; Score 31; DB 1; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1829 AAAAAAAAAAAAAAAAAAGCGCGCGC 1859
DB      43 AAAAAAAAAAAAAAAAAAGCGCGCGC 13

RESULT 8
US-08-592-820-20/c
Sequence 20, Application US/08592820
Patent No. 5759822
GENERAL INFORMATION:
APPLICANT: Chenchik, Alex
APPLICANT: Diachenko, Luda
APPLICANT: Siebert, Paul
APPLICANT: Lukianov, Sergey
APPLICANT: Gurskaya, Nadia
APPLICANT: Tarebkyin, Victor
APPLICANT: Sverdlov, Eugene
TITLE OF INVENTION: METHOD FOR SUPPRESSING DNA FRAGMENT
TITLE OF INVENTION: AMPLIFICATION DURING PCR
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESSES:
ADDRESS: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,820
FILING DATE:
CLASSIFICATION: 435
```

```

ATTORNEY/AGENT INFORMATION:
NAME: Pace, Doran R.
REGISTRATION NUMBER: 38,361
REFERENCE/DOCKET NUMBER: CL-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (synthetic)
US-08-592-820-20

Query Match      1.7%; Score 31; DB 1; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1829 AAAAAAAAAAAAAAAAAAGCGCGCGC 1859
DB      43 AAAAAAAAAAAAAAAAAAGCGCGCGC 13

RESULT 9
US-09-129-603-3/c
Sequence 3, Application US/09129603A
Patent No. 6790944
GENERAL INFORMATION:
APPLICANT: Ishiwata, Tetsuyoshi
APPLICANT: Sakurada, Mikiko
APPLICANT: Nishimura, Ayako
APPLICANT: Nakagawa, Satorshi
APPLICANT: Nishi, Tatsunari
APPLICANT: Kuga, Tetsuro
APPLICANT: No. 6790944ura, No. 6790944uo
APPLICANT: Sawada, Shigemasa
APPLICANT: Nagase, Takahiro
APPLICANT: Takei, Masami
TITLE OF INVENTION: No. 6790944el Protein
FILE REFERENCE: 766.25
CURRENT APPLICATION NUMBER: US/09/129,603A
CURRENT FILING DATE: 1998-08-05
EARLIER APPLICATION NUMBER: PCT/J97/04469
EARLIER FILING DATE: 1997-12-05
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 3
LENGTH: 50
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: other nucleic acid from homo sapiens, synthesized
US-09-129-603-3

Query Match      1.7%; Score 31; DB 4; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1829 AAAAAAAAAAAAAAAAAAGCGCGCGC 1859
DB      39 AAAAAAAAAAAAAAAAAAGCGCGCGC 9

RESULT 10
US-08-582-562A-8/c
Sequence 8, Application US/08582562A
Patent No. 5962271
GENERAL INFORMATION:
APPLICANT: Chenchik, Alex
APPLICANT: Zhu, York
```

APPLICANT: Diachenko, Lida
APPLICANT: Siebert, Paul
TITLE OF INVENTION: METHOD FOR FULL-LENGTH CDNA CLONING
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/582,562A
FILING DATE: 03-JAN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: CL-7
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (synthetic)
US-08-582-562A-8

Query Match 1.7%; Score 31; DB 2; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1829 AAAAAAAAAAAAAAAAAAGCGCGCGC 1859
DB 42 AAAAAAAAAAAAAAAAAAGCGCGCGC 12

RESULT 11
US-08-778-494B-8/C
Sequence 8, Application US/08778494B
Patent No. 5962272
GENERAL INFORMATION:
APPLICANT: Chenchik, Alex
APPLICANT: Zhu, York
APPLICANT: Diachenko, Lida
APPLICANT: Siebert, Paul
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR FULL-LENGTH CDNA
CLONING
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/778,494B
FILING DATE: 03-JAN-1997
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/582,562
FILING DATE: 03-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Doran R.
REGISTRATION NUMBER: 38,261
REFERENCE/DOCKET NUMBER: CL-7C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (352) 375-8100
TELEFAX: (352) 372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (synthetic)
US-08-778-494B-8

Query Match 1.7%; Score 31; DB 2; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1829 AAAAAAAAAAAAAAAAAAGCGCGCGC 1859
DB 42 AAAAAAAAAAAAAAAAAAGCGCGCGC 12

RESULT 12
US-08-859-998-1373/C
Sequence 1373, Application US/08859998
Patent No. 5994076
GENERAL INFORMATION:
APPLICANT: Chenchik, Alex
APPLICANT: Jokhadze, George
APPLICANT: Bibilashvili, Robert
TITLE OF INVENTION: METHOD OF ASSAYING DIFFERENTIAL
EXPRESSION
NUMBER OF SEQUENCES: 1375
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 2200 Sand Hill Road, Suite 100
CITY: Menlo Park
STATE: CA
COUNTRY: US
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/859,998
FILING DATE: 21-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Field, Bret E.
REGISTRATION NUMBER: 37,620
REFERENCE/DOCKET NUMBER: 09096/002001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-322-5070
TELEFAX: 415-854-0875
INFORMATION FOR SEQ ID NO: 1373:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-859-998-1373

Query Match 1.7%; Score 31; DB 2; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1829 AAAAAAAAAAAAAAAAAAGCGCCGC 1859
42 AAAAAAAAAAAAAAAAAAGCGCCGC 12
Db

RESULT 13
US-09-225-928-1373/C
Sequence 1373, Application US/09225928
Patent No. 6352829
GENERAL INFORMATION:
APPLICANT: Chenchik, Alex
Jokhadze, George
Bibilaashvili, Robert
TITLE OF INVENTION: METHOD OF ASSAYING DIFFERENTIAL
EXPRESSION
NUMBER OF SEQUENCES: 1375
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 2200 Sand Hill Road, Suite 100
CITY: Menlo Park
STATE: CA
COUNTRY: US
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/225,928
FILING DATE: 05-Jan-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/859,998
FILING DATE: 21-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Field, Bret E.
REGISTRATION NUMBER: 37,620
REFERENCE/DOCKET NUMBER: 09096/002001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-322-5070
TELEFAX: 415-854-0875
INFORMATION FOR SEQ ID NO: 1373:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 1373:
US-09-225-928-1373

Query Match 1.7%; Score 31; DB 3; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1829 AAAAAAAAAAAAAAAAAAGCGCCGC 1859
42 AAAAAAAAAAAAAAAAAAGCGCCGC 12
Db

RESULT 14
US-09-225-201B-1373/C
Sequence 1373, Application US/09225201B
Patent No. 6489455
GENERAL INFORMATION:
APPLICANT: Chenchik, Alex
Jokhadze, George

Bibilaashvili, Robert
TITLE OF INVENTION: METHOD OF ASSAYING DIFFERENTIAL
EXPRESSION
NUMBER OF SEQUENCES: 1375
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 2200 Sand Hill Road, Suite 100
CITY: Menlo Park
STATE: CA
COUNTRY: US
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/225,201B
FILING DATE: 05-Jan-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/859,998
FILING DATE: 21-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Field, Bret E.
REGISTRATION NUMBER: 37,620
REFERENCE/DOCKET NUMBER: 09096/002001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-322-5070
TELEFAX: 415-854-0875
INFORMATION FOR SEQ ID NO: 1373:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 1373:
US-09-225-201B-1373

Query Match 1.7%; Score 31; DB 4; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1829 AAAAAAAAAAAAAAAAAAGCGCCGC 1859
42 AAAAAAAAAAAAAAAAAAGCGCCGC 12
Db

RESULT 15
US-08-618-100B-9/C
Sequence 9, Application US/08618100B
Patent No. 6068976
GENERAL INFORMATION:
APPLICANT: Briggs, Michael R.
APPLICANT: Auwerx, Johan
APPLICANT: de Vos, Piet
APPLICANT: Staelens, Bart
APPLICANT: Croston, Glenn E.
APPLICANT: Miller, Stephen G.
TITLE OF INVENTION: MODULATORS OF OB GENE AND
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb


```

MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/618,100B
FILING DATE: March 19, 1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/558,588
FILING DATE: October 30, 1995
APPLICATION NUMBER: 08/510,584
FILING DATE: August 2, 1995
APPLICATION NUMBER: 08/418,096
FILING DATE: April 5, 1995
APPLICATION NUMBER: 08/408,584
FILING DATE: March 20, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 219/075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ. ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 52 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
OTHER INFORMATION: "N" represents any base.
US-08-618-100B-9

```

```

Query Match      1.7%; Score 31; DB 3; Length 52;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1829 AAAAAAAAAAAAAAAAAAGCGCGCCG 1859
Db      43 AAAAAAAAAAAAAAAAAAGCGCGCCG 13

```

Search completed: December 30, 2004, 03:23:40
Job time : 213 secs

THIS PAGE IS BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2004, 20:34:24 ; Search time 8793 Seconds
(without alignments)
7704.022 Million cell updates/sec

Title: US-09-993-966-5

Sequence: 1 gaattcgcccttctaactg.....aaaaaaaaaagcgcgccgc 1859

Scoring table: (OLIGO NUC)
Gapop 60.0, Gapext 60.0

Searched: 32822875 seqs, 18219865908 residues

Word size: 0

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_hic: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gsa1: *
9: gb_gsa2: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1525	82.0	2142	3 AF289584	AF289584 Homo sapi
2	885	47.6	885	9 AY412098	AY412098 Homo sapi
3	814	43.8	931	5 BQ653673	BQ653673 AGENCOURT
4	748	40.2	921	5 BQ645656	BQ645656 AGENCOURT
5	702	37.8	953	5 BU846054	BU846054 AGENCOURT
6	689	37.1	990	5 BQ64678	BQ64678 AGENCOURT
7	660	35.5	895	5 BQ644360	BQ644360 AGENCOURT
8	659	35.4	928	5 BQ644956	BQ644956 AGENCOURT
9	640	34.4	996	5 BQ652087	BQ652087 AGENCOURT
10	639	34.4	966	5 BQ649813	BQ649813 AGENCOURT
11	622	32.5	822	6 CB215756	CB215756 NISC np09
12	610	32.8	825	6 CB961961	CB961961 AGENCOURT
13	593	31.9	634	7 CN369119	CN369119 AGENCOURT
14	582	31.3	627	4 B1767278	B1767278 603057995
15	574	30.9	1027	4 BQ104777	BQ104777 603311726
16	556	29.9	1008	5 BQ646371	BQ646371 AGENCOURT
17	551	29.6	826	6 CB961668	CB961668 AGENCOURT
18	539	29.0	1059	4 BQ070932	BQ070932 AGENCOURT
19	508	27.3	364	4 BM711145	BM711145 UI-E-DX1-
20	465	25.0	899	5 BQ645507	BQ645507 AGENCOURT
21	452	24.3	885	9 AY412099	AY412099 Pan trogl
22	450	24.2	751	4 BG820139	BG820139 602782356
23	446	24.0	931	5 BU856404	BU856404 AGENCOURT
24	429	23.1	481	4 B1047069	B1047069 MR3-FM020

25	397	21.4	777	4 BG542261	BG542261 602571809
26	395	21.2	395	1 A1167910	A1167910 OK29b05.X
27	334	18.0	429	9 CN369118	CN369118 I70005322
28	326	17.5	654	9 AG112447	AG112447 Pan trogl
29	244	13.1	333	2 BF436193	BF436193 nab45a02.
30	220	11.8	890	5 BQ645220	BQ645220 AGENCOURT
31	220	11.8	959	5 BQ653014	BQ653014 AGENCOURT
32	209	11.2	565	2 BF920913	BF920913 MR3-NT013
33	188	10.1	500	4 BM670652	BM670652 UI-E-DX1-
34	126	6.8	638	2 AG053170	AG053170 Pan trogl
35	104	5.6	550	2 AW752389	AW752389 RC2-C7020
36	78	4.2	555	5 BP446790	BP446790 BP446790
37	70	3.8	684	7 CK450648	CK450648 903739 MA
38	61	3.3	252	8 B91703	B91703 RBC111-20A1
39	60	3.2	1487	6 CB381772	CB381772 UZ11M15.2
40	56	3.0	661	1 AU260449	AU260449 AU260449
41	55	3.0	631	5 BQ622880	BQ622880 ZOCR24 C
42	54	2.9	214	5 BQ622887	BQ622887 ZOCR24 C
43	54	2.9	254	5 BQ622885	BQ622885 ZOCR29 C
44	54	2.9	639	5 BQ622909	BQ622909 ZOCR29 C
45	54	2.9	707	5 BQ622883	BQ622883 ZOCR27 C

ALIGNMENTS

RESULT 1	AF289584	2142 bp	mRNA	linear	HTC 01-JAN-2002
LOCUS	Homo sapiens clone p7246	unknown	MRNA.		
DEFINITION	AF289584				
ACCESSION	AF289584.1	GI:18027371			
VERSION					
KEYWORDS	HTC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 2142) Zhang, P.P., Zhou, X.M., Jiang, H.Q., Huang, Y., Qin, W.X., Zhao, X.T., Wan, D.F. and Gu, J.R. Novel human cDNA clones with function of inhibiting cancer cell growth				
TITLE	Unpublished				
JOURNAL	2 (bases 1 to 2142) Zhang, P.P., Zhou, X.M., Jiang, H.Q., Huang, Y., Qin, W.X., Zhao, X.T., Wan, D.F. and Gu, J.R. Direct Submission				
REFERENCE	Submitted (17-JUL-2000) National Laboratory for Oncogenes & Related Genes, Shanghai Cancer Institute, 25/Jun 2200 Xie-Tu Road, Shanghai 200032, P. R. China				
FEATURES	Location/Qualifiers				
source	1..2142				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="p7246"				
	277..1689				
	/product="unknown"				
	/protein_id="FAL55768.1"				
	/db_xref="GI:18027372"				
	/translation="MGLSKPYDVCKRRESPGDSFVNSAARKIEIEMTRORCP GGVSGRQLACTIGRRELVDGLVRLTSEEDDFLEVALPEPKTDGASGDG KKMERSEPCPGSKQIKREBLQCDVSMEDSQEWTFLLYDNNKXVREDITSL HTIYEVDSSVNSPTSSKMLRYKLTVAIDGSDQSRSVNVQADLSARPAETPTE DLRSMKKORAPLRFQDGRLEBSGCHHCVDNIRRNHYDLAGIENTSGFPGS PSYAKSELPPRTSNPTRSRSHPEAIIHPKPDGVDPASFFRLDPLAKVSELOR LRGTQDSKHTFVRSFPAQGSYGVGVAVAGARKRPPLGPAIPVSSAHLAASPLP SLAPLGHKKHKKAKESQGCRCGLQAPLVSAGGVLDGRHLEPLALVYESQAGQPVQ RHEHHHHHHHHHHFYQT"				
ORIGIN	Query Match	82.0%	Score 1525;	DB 3;	Length 2142;

Best Local Similarity 99.9%; Pred. No. 0;
Matches 1575; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	253	CGTGTGCAAGCGCAGGAGAGCCCGGAAAGGTGACAGCTTCCGCTGTGAGCGCTGCTGGGC	312
Db	306	CGTGTGCAAGCGCAGGAGAGCCCGGAAAGGTGACAGCTTCCGCTGTGAGCGCTGCTGGGC	365
QY	313	TCGGAAGGGCAATCGAGAGTGGATCGGGAGACAGCGCTGCCGGGCGGTGTCTCGGAGCC	372
Db	366	TCGGAAGGGCAATCGAGAGTGGATCGGGAGACAGCGCTGCCGGGCGGTGTCTCGGAGCC	425
QY	373	CCGACAGCTCGGTTGCGGGCACCATAAGCCGGAACACCCGGAGCTCGTGGCGACGT	432
Db	426	CCGACAGCTCGGTTGCGGGCACCATAAGCCGGAACACCCGGAGCTCGTGGCGACGT	485
QY	433	GTTGAGAGACACGCTCAACCGAGGAAGAGAGACGACTTTGGGCTGTGAAAGTGGCCCTGCC	492
Db	486	GTTGAGAGACACGCTCAACCGAGGAAGAGAGACGACTTTGGGCTGTGAAAGTGGCCCTGCC	545
QY	493	TCCTGAGAAAGCTGACGGGCTGGGCAAGCGAGATGAGAAAGATGAGAGAGTGAAGCA	552
Db	546	TCCTGAGAAAGCTGACGGGCTGGGCAAGCGAGATGAGAAAGATGAGAGAGTGAAGCA	605
QY	553	ACCTTGCCCAAGCTTCCAAAGCAGCTGAAAGTTTGAAGAGTCCAGTGCAGCGTGTCCAT	612
Db	606	ACCTTGCCCAAGCTTCCAAAGCAGCTGAAAGTTTGAAGAGTCCAGTGCAGCGTGTCCAT	665
QY	613	GGAAGAGACAGCCGGGACGGAGTGAACCTTCAACCTGTATGACCTTTGACAAACAAGGCA	672
Db	666	GGAAGAGACAGCCGGGACGGAGTGAACCTTCAACCTGTATGACCTTTGACAAACAAGGCA	725
QY	673	GATCAACCCGAGAGACATCACAGCTTGTCTGACACACATCTATGAGGTGTGACTCTCTC	732
Db	726	GATCAACCCGAGAGACATCACAGCTTGTCTGACACACATCTATGAGGTGTGACTCTCTC	785
QY	733	TGTCAACCACTCCCAACATCCAGCAGATGCTGGGGTAAAGCTCACCGTGGCCCCGA	792
Db	786	TGTCAACCACTCCCAACATCCAGCAGATGCTGGGGTAAAGCTCACCGTGGCCCCGA	845
QY	793	TGGGACGCCAGAGCAAGAGAGACGTCCTTGTCAATACAGGCTGACCTGACAGACCGCAAGGCC	852
Db	846	TGGGACGCCAGAGCAAGAGAGACGTCCTTGTCAATACAGGCTGACCTGACAGACCGCAAGGCC	905
QY	853	CCGAGCAGAGACCAAGCCCACTGAGAGACCTGCGGAGCTGTGGAGAAAGACAGCAGCCCC	912
Db	906	CCGAGCAGAGACCAAGCCCACTGAGAGACCTGCGGAGCTGTGGAGAAAGACAGCAGCCCC	965
QY	913	GCTCAGGTTCCAGGGTGAACAGCCGCTTGAGACAGTCTGGCTGTACCAACAATTGCGTGA	972
Db	966	GCTCAGGTTCCAGGGTGAACAGCCGCTTGAGACAGTCTGGCTGTACCAACAATTGCGTGA	1025
QY	973	TGAGAACATGAGAGAGAAACCACTATTGATCTCGCCGGGATGAGAAACTACACGTC	1032
Db	1026	TGAGAACATGAGAGAGAAACCACTATTGATCTCGCCGGGATGAGAAACTACACGTC	1085
QY	1033	CCAATTTGGGCTCGGCTCCCTTCCGTGGGCCAGAGATCAAGACCTGCCCCCGCACCTC	1092
Db	1086	CCAATTTGGGCTCGGCTCCCTTCCGTGGGCCAGAGATCAAGACCTGCCCCCGCACCTC	1145
QY	1093	CAATTCACCTGATCTCGCTCCCATGAGCCGGAAGCATTCACATCCACAACGAAAGCC	1152
Db	1146	CAATTCACCTGATCTCGCTCCCATGAGCCGGAAGCATTCACATCCACAACGAAAGCC	1205
QY	1153	CCAAAGGGGTGACCCGGGCTCTTCCATTCCTTTGACACCCCAATTCGCAAGGGTCTCAGA	1212
Db	1206	CCAAAGGGGTGACCCGGGCTCTTCCATTCCTTTGACACCCCAATTCGCAAGGGTCTCAGA	1265
QY	1213	GCCTCAGCAACGGCTCCGGGGCACCCAGAGACGGGACCAAGCATTTTGTGAGGTCCCCCA	1272
Db	1266	GCCTCAGCAACGGCTCCGGGGCACCCAGAGACGGGACCAAGCATTTTGTGAGGTCCCCCA	1325
QY	1273	GGCCCAAGGCAAGAGTGTGGTGTGGGCCACGTGGCCAGAGGGGCAAGAAACAAGCCCC	1332

Db	1326	GGCCGAGGGAAGAAGTGTGGGTGTGGGCACAGTGGCCAGAGGGGCAAGAAAACAAGCCCC	1385
QY	1333	TCTGGGACCCGCCACATCCCTGCGGTGTCCCCCTCCGCCACCTGTGCTCCAGCCCGGCGCT	1392
Db	1386	TCTGGGACCCGCGCATCCCTGGGTGTCCCCCTCGGCCACCTGGCTGCAAGCCGGCGCT	1445
QY	1393	CCTCCCTCCTTAAGCCCCCTCGGGCACAGAGCACAGCATCGAGCCAAAGAGAGGCCA	1452
Db	1446	CCTCCCTCCTTAAGCCCCCTCGGGCACAGAGCACAGCATCGAGCCAAAGAGAGGCCA	1505
QY	1453	GAAGGGCTGGCGGGGCGTGCAGAGGACCATGAGGCTCATGGTGGCCCTGACCGGGGCGGGA	1512
Db	1506	GAAGGGCTGGCGGGGCGTGCAGAGGACCATGAGGCTCATGGTGGCCCTGACCGGGGCGGGA	1565
QY	1513	GCACCTGGGGAGCTGCGCCGCTTGTTGTGTGTATGAGAGCCAGGCGGGGACGCCGTCCA	1572
Db	1566	GCACTGGGGAGGCTGCGCCGCTTGTTGTGTGTATGAGAGCCAGGCGGGGACGCCGTCCA	1625
QY	1573	GAGACATGAGCACACCAACCATGAAATCACCAATTAACCAACATTTCACGAC	1632
Db	1626	GAGACATGAGCACACCAACCATGAAATCACCAATTAACCAACATTTCACGAC	1685
QY	1633	ATAAGGCCCCCCCCAGGGCCCCCACTGTGCATATGAAGACCCCAACCCCGACACACA	1692
Db	1686	ATAAGGCCCCCCCCAGGGCCCCCACTGTGCATATGAAGACCCCAACCCCGACACACA	1745
QY	1693	AGGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1752
Db	1746	AGGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1805
QY	1753	TTACTCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1812
Db	1806	TTACTCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1865
QY	1813	CTTTATTTATTAATGTT	1828
Db	1866	CTTTATTTATTAATGTT	1881

RESULT 2						
AY412098	AY412098	885 bp	DNA	linear	GSS 16-DEC-2003	
LOCUS	Homo sapiens NKD1 gene, VIRUAL TRANSCRIPT, partial sequence,					
DEFINITION	Genomic survey sequence.					
ACCESSION	AY412098					
VERSION	AY412098.1	GI:39768063				
KEYWORDS	GSS.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					

REFERENCE

1 (bases 1 to 885)

AUTHORS

Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A.

1000, M.A., Tallentire, D.M., Caveiro, D.A., Lee, E., Murphy, J., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.

JOURNAL
Science 302 (5652) : 1960-1963 (2003)

RECEIVED	146/1302
AUTHORS	2 (bases 1 to 885)
REFERENCE	Clark A.G., Glanowski, S., Nielson, R., Thomas, P., Ketariwal, A.

Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
 Ferreria, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J.,
 Adams, M.D., and Carroll, M.

TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) *Celera Genomics*, 45 West Gude Drive,
Rockville, MD 20850, USA

FEATURES	COMMENT
Location/Qualifiers	This sequence was made by sequencing genomic exons and ordering them based on alignment.

```
source      1. .885
/organism="Homo sapiens"
```


QY 1182 TCCTTGAACCCCAATGCGCAAGGCTCAGAGCTCCAGCAACGGCTCCGGGGCACCAGG 1241
DB 361 TCCTTGAACCCCAATGCGCAAGGCTCAGAGCTCCAGCAACGGCTCCGGGGCACCAGG 420
QY 1242 ACCGGAACAAGCACTTTGTAGAGTCCGCCAAGGCCCAAGGCAAGTGTGGGTGTGGCC 1301
DB 421 ACCGGAACAAGCACTTTGTAGAGTCCGCCAAGGCCCAAGGCAAGTGTGGGTGTGGCC 480
QY 1302 ACGTGGCCAGAGGGGCGAAGAAACAAGCCCTCTGGGAACTCCGATCTTCGGGTGTCCC 1361
DB 481 ACGTGGCCAGAGGGGCGAAGAAACAAGCCCTCTGGGAACTCCGATCTTCGGGTGTCCC 540
QY 1362 CCTCGGCGCACTGGCTGGCCAGCGCGGCTCTCCCTCCCTGATGCCCGCTCGGGACA 1421
DB 541 CCTCGGCGCACTGGCTGGCCAGCGCGGCTCTCCCTCCCTGATGCCCGCTCGGGACA 600
QY 1422 AGAAGCAACAAGCAACGAGCCAGAGGAGAGCCAGAGGCTGCGGGGCTTGACGACAC 1481
DB 601 AGAAGCAACAAGCAACGAGCCAGAGGAGAGCCAGAGGCTGCGGGGCTTGACGACAC 660
QY 1482 TGGGCTCAGGTGGCTCTGTCTTGGGGGCGGAGCACTTGGGAGCTGCGGCTTGGTGG 1541
DB 661 TGGGCTCAGGTGGCTCTGTCTTGGGGGCGGAGCACTTGGGAGCTGCGGCTTGGTGG 720
QY 1542 TGTATGAGAGCGAGGCGGGGAGCGGGTCCAGAGACATGAGCACACACCATGTAAC 1601
DB 721 TGTATGAGAGCGAGGCGGGGAGCGGGTCCAGAGACATGAGCACACACCATGTAAC 780
QY 1602 ATCACCAACATTACCACTTCTTACGACATA 1635
DB 781 ATCACCAACATTACCACTTCTTACGACATA 814
RESULT 4
LOCUS B0645656 921 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT 8355700 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6285268
5', mRNA sequence.
ACCESSION B0645656
VERSION B0645656.1 GI:21769828
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 921)
TITLE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LNCM2484 row: d column: 05
High quality sequence start: 16
High quality sequence stop: 685.
Location/Qualifiers
1..921
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6285268"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_100"
/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using Zap-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

ORIGIN

Query Match 40.2%; Score 748; DB 5; Length 921;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 748; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 825 ATCAGCTGACCTGACAGAGGCAAGGCGCCGAGCAAGACCAAGCCACTGAGGACTGCG 884
DB 23 ATCAGCTGACCTGACAGAGGCAAGGCGCCGAGCAAGACCAAGCCACTGAGGACTGCG 82
QY 885 GAGAGCTGGAGAAAGAAAGCAAGCAGAGCCCGCTCAGGTTCCAGGGTGAAGCCGCTGGAGC 944
DB 83 GAGAGCTGGAGAAAGAAAGCAAGCAGAGCCCGCTCAGGTTCCAGGGTGAAGCCGCTGGAGC 142
QY 945 AGTCTGCTCTTACCACTATTGGGTATGAGAACTGAGAGAGAAACCACTACTTAG 1004
DB 143 AGTCTGCTCTTACCACTATTGGGTATGAGAACTGAGAGAGAAACCACTACTTAG 202
QY 1005 ATCTCGCGGGATAGAAAATTACACGTCCTCAATTTTGGGCTGGCTCCCTCCGTGGCC 1064
DB 203 ATCTCGCGGGATAGAAAATTACACGTCCTCAATTTTGGGCTGGCTCCCTCCGTGGCC 262
QY 1065 AGAAGTCAGAACTGCCCCCGGACCTTCCATTCACCTGATTCCTCCATGAGCCGG 1124
DB 263 AGAAGTCAGAACTGCCCCCGGACCTTCCATTCACCTGATTCCTCCATGAGCCGG 322
QY 1125 AAGCCATTCACATTCACACCCGAAAGCCCGGCTGGACCCGGCTCTTCCACTTCC 1184
DB 323 AAGCCATTCACATTCACACCCGAAAGCCCGGCTGGACCCGGCTCTTCCACTTCC 382
QY 1185 TTGACACCCCAATTCGCAAGGCTCTCAGAGCTCCAGCAACGGCTCCGGGGCACCAGGAGC 1244
DB 383 TTGACACCCCAATTCGCAAGGCTCTCAGAGCTCCAGCAACGGCTCCGGGGCACCAGGAGC 442
QY 1245 GAGCAAGCACTTGTAGAGTCCGCCAAGGCCCAAGGCGCAAGAGTGTGGTGGCCAGC 1304
DB 443 GAGCAAGCACTTGTAGAGTCCGCCAAGGCCCAAGGCGCAAGAGTGTGGTGGCCAGC 502
QY 1305 TGGCCAGAGGGGCAAGAAACAAGCCCTCTGGGACCCCGCATCCCTGGCGGTGCCCCCT 1364
DB 503 TGGCCAGAGGGGCAAGAAACAAGCCCTCTGGGACCCCGCATCCCTGGCGGTGCCCCCT 562
QY 1365 CCGGCCACCTGGCTGCGAGGCCGCGGCTCTCCCTCCCTCCTAGGCCCTCGGGCAACAAG 1424
DB 563 CCGGCCACCTGGCTGCGAGGCCGCGGCTCTCCCTCCCTCCTAGGCCCTCGGGCAACAAG 622
QY 1425 AGCAACAAGCACGAGCCCAAGAGAGACCAAGAGGCTGCGGGGCTTGAGAGCACTGG 1484
DB 623 AGCAACAAGCACGAGCCCAAGAGAGACCAAGAGGCTGCGGGGCTTGAGAGCACTGG 682
QY 1485 CTTGAGGTGGCTCTGTCTTGGGGCGGAGACACTTGGGGAGCTGCGGCTTGGTGTGT 1544
DB 683 CTTGAGGTGGCTCTGTCTTGGGGCGGAGACACTTGGGGAGCTGCGGCTTGGTGTGT 742
QY 1545 ATGAGAGCCAGGCGGGGACGCGGTCA 1572
DB 743 ATGAGAGCCAGGCGGGGACGCGGTCA 770
RESULT 5
LOCUS BU846054 953 bp mRNA linear EST 16-OCT-2002
DEFINITION AGENCOURT 10413301 NIH_MGC_109 Homo sapiens cDNA clone
IMAGE:6580002 5', mRNA sequence.
ACCESSION BU846054
VERSION BU846054.1 GI:24030759

KEYWORDS EST. Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 953)

TITLE NIH-MGC http://mgc.nci.nih.gov/

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strauberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

Plate: LCM2784 row: d column: 18

High quality sequence stop: 627.

Location/Qualifiers

1..953

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6580002"

/issue_type="teratocarcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH MGC 109"

/note="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2: XhoI; CDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC library."

ORIGIN

Query Match 37.8%; Score 702; DB 5; Length 953;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 702; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 AGTCGGGCGCGCGGACGCGCGGACGAGCGCGTCCGCGCGCGCTCGGCGCTCGCG 125

DB 66 AGTCGGGCGCGCGGACGCGCGGACGAGCGCGTCCGCGCGCGCTCGGCGCTCGCG 125

QY 126 CTGCGGGGGCGCTTCGGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 165

DB 126 CTGCGGGGGCGCTTCGGGAGGAGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 185

QY 186 CATGGCTTAAAGGAGCTCCCGCGCGCGGACGCCCATGGGAACTTCACTCCAGC 245

DB 186 CATGGCTTAAAGGAGCTCCCGCGCGCGGACGCCCATGGGAACTTCACTCCAGC 245

QY 246 CGGCGCGCGTGTGCAAGCGAGGAGAGCCCGGAGGTGACAGCTTCCGCTGAGCGCTG 305

DB 246 CGGCGCGCGTGTGCAAGCGAGGAGAGCCCGGAGGTGACAGCTTCCGCTGAGCGCTG 305

QY 306 CCGGGGCTCGGAAGGAGGATGAGAGATGATTCGGGAGACAGCGCTGCCCGGCGGTCT 365

DB 306 CCGGGGCTCGGAAGGAGGATGAGAGATGATTCGGGAGACAGCGCTGCCCGGCGGTCT 365

QY 366 CGGAGACCCGACAGCTGCGGTGGCGGGACCAATAGGCCAAGCACCCGGAGAGCTCGTG 425

DB 366 CGGAGACCCGACAGCTGCGGTGGCGGGACCAATAGGCCAAGCACCCGGAGAGCTCGTG 425

QY 426 GCGACGTGTTGAGAGACAGCTCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485

DB 426 GCGACGTGTTGAGAGACAGCTCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485

QY 486 CCGTGCCTCTGAG 545

DB 486 CCGTGCCTCTGAG 545

DB 486 CCGTGCCTCTGAG 545

QY 546 TGAGCCAAACCTTGCCCAAG 605

DB 546 TGAGCCAAACCTTGCCCAAG 605

QY 606 TGTCATGAG 665

DB 606 TGTCATGAG 665

QY 666 ACCGCAAGGTGACCCGAG 725

DB 666 ACCGCAAGGTGACCCGAG 725

QY 726 ACTCTCTGTGCAACCACTCCCAACATCCAGCAAGATGCTGC 767

DB 726 ACTCTCTGTGCAACCACTCCCAACATCCAGCAAGATGCTGC 767

RESULT 6

BQ064678

LOCUS BQ064678

DEFINITION AGENCOURT 6853565 NIH_MGC_99 Homo sapiens CDNA clone IMAGE:5926427

5', mRNA Sequence.

ACCESSION BQ064678

VERSION BQ064678.1 GI:19893537

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NIH-MGC http://mgc.nci.nih.gov/

TITLE 1 (bases 1 to 960)

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strauberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Lon Staudt
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

Plate: LCM2099 row: 1 column: 12

High quality sequence stop: 670.

Location/Qualifiers

1..960

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5926427"

/issue_type="lymphoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_99"

/note="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; CDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC library."

ORIGIN

Query Match 37.1%; Score 689; DB 5; Length 960;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 739; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 822 TCATCAGGCTGACCTGACAGAGCGCAAGGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 881

DB 822 TCATCAGGCTGACCTGACAGAGCGCAAGGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 881

Db		199	ATCTCGCGGGGATAGAAAACTAACA	GTCCTCAATTGTGGGCCTGGCTCCCTTCCGCGGCC	258
QY		1065	AGAAGTCAGAAATGCTCCCCCGGCAC	TCTCCTCAATCCCATCTCGATCTTCGCTCCATAGCCGG	1122
Db		259	AGAAGTACGAATGCTCCCCCGGCAC	TCTCCTCAATCCCATCTCGATCTTCGCTCCATAGCCGG	318
QY		1125	AAGCCATTCACATCCGACACCGA	AAAGCCCAGAGCGTGGAACCCGAGCCTCTTCCACTTCC	1184
Db		319	AAGCCATTCACATCCGACACCGA	AAAGCCCAGAGCGTGGAACCCGAGCCTCTTCCACTTCC	378
QY		1185	TTGACACCCCAATCGCCCAAGGTCT	CAGAGCTCCAGCAA	CGGCTTCGCGGACACCAAGACG 1244
Db		379	TTGACACCCCAATCGCCCAAGGTCT	CAGAGCTCCAGCAA	CGGCTTCGCGGACACCAAGACG 438
QY		1245	GGAGCAGCACCTTTGTAGAGTCCCC	CCAAGGCCCAAGGCGCAGAGTGTGGGTGTGGGCCACG	1304
Db		439	GGAGCAGCACCTTTGTAGAGTCCCC	CCAAGGCCCAAGGCGCAGAGTGTGGGTGTGGGCCACG	498
QY		1305	TGGCCAAAGGGGGCAGAAACAAG	CCCCCTCTGGGAGCCCGCATCCCTGCGGTGTCCCTCT	1364
Db		499	TGGCCAAAGGGGGCAGAAACAAG	CCCCCTCTGGGAGCCCGCATCCCTGCGGTGTCCCTCT	558
QY		1365	CCGCCCCACTGTGCTGCTGCGAG	CCCGGCGCTCTCTCCCTCTTAGACC	CCCCCTCGGGGACACAAGA 1422
Db		559	CCGCCCCACTGTGCTGCTGCGAG	CCCGGCGCTCTCTCCCTCTTAGACC	CCCCCTCGGGGACACAAGA 618
QY		1425	AGCACAAGCACCGAGCCCAAGAGA	GCAGCAGCAGGCTGCGGGGCTCTGCAGGACACCACTGG	1484
Db		619	AGCACAAGCACCGAGCCCAAGAGA	GCAGCAGCAGGCTGCGGGGCTCTGCAGGACACCACTGG	678
QY		1485	CCTCAGGTGGCCCTGTCTCTGGG	CGGGAGACACCTGCGGAGCTCCCGGCC	1534
Db		679	CCTCAGGTGGCCCTGTCTCTGGG	CGGGAGACACCTGCGGAGCTCCCGGCC	728
RESULT_9					
LOCUS	B0652087				
DEFINITION	AGEBNCOURT_8207495 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6283110				
ACCESSION	B0652087				
VERSION	B0652087.1				
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.				
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Straubeberg, Ph.D. Email: ggaabs-remail.nih.gov Tissue Procurement: CGAP (Stanford) CDNA Library Preparation: Rubin Laboratory DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) Genome Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.nhl.gov Plate: LUCW2478 row: j column: 07 High quality sequence stop: 604. Location/Qualifiers 1..996 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /feature="IMAGE:6283110" /tissue_type="hepatocellular carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH MGC 100" /note="Organ: liver; Vector: pOTB1; site_1: XhoI; site_2:				
FEATURES					
SOURCE					

EcotRI; cDNA made by oligo-dT priming. Directionally cloned into EcotRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."

ORIGIN

Query Match 34.4%; Score 640; DB 5; Length 966;
 Best Local Similarity 99.9%; Pred. No. 1.3e-292;
 Matches 690; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

829 GGCTGACCTGACGAGCGCAAGGCCCCGAGAGAGACCAACCACTGAGACCTGCGGAG 888
 8 GGCTGACCTGACGAGCGCAAGGCCCCGAGAGAGACCAACCACTGAGAGCTGCGGAG 67

889 CTGGAG 948
 68 CTGGAG 127

949 TGCGTCTACCACTATGCGTAGATGAGACATGAGAGAGAGAGAGAGAGAGAGAG 1008
 128 TGCGTCTACCACTATGCGTAGATGAGACATGAGAGAGAGAGAGAGAGAGAGAG 187

1009 CGCGGAGATGAG 1068
 188 CGCGGAGATGAG 247

1069 GTGAG 1128
 248 GTGAG 307

1129 CATCCACATCCACACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1188
 308 CATCCACATCCACACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 367

1189 CACCCCAATGCGCAAGGCTTCAGAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAG 1248
 368 CACCCCAATGCGCAAGGCTTCAGAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAG 427

1249 CAAGACATTTGTGAGTCCCCCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1308
 428 CAAGACATTTGTGAGTCCCCCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 487

1309 CAGAGGAG 1368
 488 CAGAGGAG 547

1369 CCACCTGGCTGCGAGCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1428
 548 CCACCTGGCTGCGAGCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 607

1429 CAAGACATTTGTGAGTCCCCCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1488
 608 CAAGACATTTGTGAGTCCCCCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 667

1489 AGGTGGCCCTGTCTCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1519
 668 AGGTGGCCCTGTCTCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 698

RESULT 10
 B0649813
 LOCUS AGNCOURT 8302499 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6271026
 DEFINITION 5', mRNA sequence.
 ACCESSION B0649813
 VERSION B0649813.1 GI:21773985
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>,
 1 (bases 1 to 966)
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: sgabs-r@mail.nih.gov
 Tissue Procurement: CGAP (Stanford)

CDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: LICR2447 row: b column: 19
 High quality sequence stop: 583.
 Location/Qualifiers

FEATURES

1..966
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6271026"
 /issue_type="hepatocellular carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 100"
 /note="Organ: Liver; Vector: pORF7; Site 1: XhoI; Site 2: EcotRI; cDNA made by oligo-dT priming. Directionally cloned into EcotRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."

ORIGIN

Query Match 34.4%; Score 639; DB 5; Length 966;
 Best Local Similarity 99.9%; Pred. No. 3.9e-292;
 Matches 689; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

825 ATGAGCTGACCTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 884
 3 ATGAGCTGACCTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 62

885 GAGAGCTGGAG 944
 63 GAGAGCTGGAG 122

945 AGTCTGGCTCTCCACCATTTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1004
 123 AGTCTGGCTCTCCACCATTTGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 182

1005 ATCTCGCGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1064
 183 ATCTCGCGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 242

1065 AGAAGTCAGAACTGCCCCCGCACTCCATCTCGATCTCGTCCCATGAGCGG 1124
 243 AGAAGTCAGAACTGCCCCCGCACTCCATCTCGATCTCGTCCCATGAGCGG 302

1125 AAGCCATCCACATCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1184
 303 AAGCCATCCACATCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 362

1185 TTGACACCCCAATGCGCAAGGCTTCAGAGCTTCAGAGAGAGAGAGAGAGAGAG 1244
 363 TTGACACCCCAATGCGCAAGGCTTCAGAGCTTCAGAGAGAGAGAGAGAGAGAG 422

1245 GAG 1304
 423 GAG 482

Oy		1305	TGGCAGAGGAGGGAAGAAACAAGCCCTCTGTGGGACCGGCATCCGTCGGTGTCCCTC	1364
Dd		483	TGGCCAGAGGGGCAAGAAACAAGCCCTCTGTGGGACCGGCATCCGTCGGTGTCCCTC	542
Oy		1365	CCGCCCACTTGCTGTCAGACCCCGGCTCTCTCCCTCTAGCCCCCTCGGGACAAGA	1424
Dd		543	CCGCCCACTTGCTGTCAGACCCCGGCTCTCTCCCTCTAGCCCCCTCGGGACAAGA	602
Oy		1425	AGCACAGACACCAGACCCAAGAGACCGACGGCTGCGGGGCTGTGAGGACCACTGG	1484
Dd		603	AGCACAGACACCAGACCCAAGAGACCGACGGCTGCGGGGCTGTGAGGACCACTGG	662
Oy		1485	CCTCAGGTGGCCCTGTCTCTGGGGGGGAGC	1514
Dd		663	CCTCAGGTGGCCCTGTCTCTGGGGGGGAGC	692
RESULT_11				
CB215756				
LOCUS				
DEFINITION		CB215756	622 bp	mRNA linear EST 06-FEB-2003
ACCESSION		NISC_np09e12.y1 NICHD_HS_Ut1	Homo sapiens	cDNA clone IMAGE:5937502
VERSION		CB215756		5', mRNA sequence.
KEYWORDS		CB215756.1	GI:28263948	EST..
SOURCE				
ORGANISM				Homo sapiens (human)
REFERENCE				Homo sapiens
AUTHORS				Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates, Catarrhini; Hominidae; Homo. NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov cDNA Library Preparation: CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LNL DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC) Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: info@image.lnl.gov Plate: LHAM3164 row: I column: 23 Seq primer: M13Rpl reverse primer (AB1). Location/Qualifiers 1..622 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:5937502" /sex="female" /tissue_type="normal endometrium, late proliferative phase, cycle day 13" /lab_host="DH10B (TI-resistant)" /clone_id="NICHD_HS_Ut1" /note="Organ: uterus; Vector: pCMV-SPORT6.1.ccdB (Resgen, Invitrogen Corporation); Site 1: NotI; Site 2: EcoRV; Cloned unidirectionally from microquantity amounts of mRNA from normal endometrial tissue (late proliferative phase, cycle day 13). Average insert size 1.9 kb. Library constructed by Resgen (Invitrogen Corporation)."
FEATURES				
source				
ORIGIN				
Query Match		33.5%	Score 622;	DB 6; Length 622;
Best Local Similarity		100.0%;	Pred. No. 4.8e-264;	
Matches 622;		Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Oy		1159	CGTGAACCCGGCCTTCCTTCACTTTGACACCCCATCGCAAGTCTTCAGAGCTCCA	1218
Dd		1	CGTGAACCCGGCCTTCCTTCACTTTGACACCCCATCGCAAGTCTTCAGAGCTCCA	60
Oy		1219	GCAACGGCTCCGGGAGACCCAGACGGGAGACAGCATTTGTAGAGTCCCACCAAGCCCCA	1278

[illegible]

```
/clone="IMAGE:30349111"
/tissue_type="pre-eclampitic placenta"
/lab_host="DH10B Tona"
/clone_id="NIH_MGC_148"
/note="Organ: placenta; Vector: pBluescriptR; Site 1:
all-XhoI; Site 2: BamH; Library is oligo-dT primed and
directionally cloned using primer
5'-TTTTTTTTTTTTTNN-3', size selected for average insert
size 2.3 kb and normalized to R0.5. This is a primary
library enriched for full-length clones and constructed
using the Cap-trapper method (Carninci, in preparation).
Library constructed by M. Brownstein (NIMH/NHGRI,
National Institutes of Health). Note: this is a NIH_MGC
Library."
```

ORIGIN

Query Match 32.8%; Score 610; DB 6; Length 825;
Best Local Similarity 100.0%; Pred. No. 2,4e-278;
Matches 610; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 65 GAGTCGGGCGCGCGGCGAGAGCGCGTCCCGCGCGCGCTCGGCTCCGCTCG 124
DB 27 GAGTCGGGCGCGCGGCGAGAGCGCGTCCCGCGCGCGCTCGGCTCCGCTCG 86
QY 125 GCTCGGGGCTGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 184
DB 87 GCTCGGGGCTGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 146
QY 185 GCATGCTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 244
DB 147 GCATGCTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 206
QY 245 CCGGCGCGCGCGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 304
DB 207 CCGGCGCGCGCGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 266
QY 305 GCTCGGGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 364
DB 267 GCTCGGGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 326
QY 365 TCGGAGACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 424
DB 327 TCGGAGACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 386
QY 425 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 484
DB 387 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 446
QY 485 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 544
DB 447 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 506
QY 545 GTGAGCGAACCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 604
DB 507 GTGAGCGAACCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 566
QY 605 GTGTCCTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 664
DB 567 GTGTCCTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 626
QY 665 AACGGCAAGG 674
DB 627 AACGGCAAGG 636
```

RESULT 13
CN369119 634 bp mRNA linear EST 16-MAY-2004
LOCUS CN369119
DEFINITION 1700059936909 GRN_PRENBU Homo sapiens cDNA 5', mRNA sequence.
ACCESSION CN369119
VERSION CN369119.1 GI:47369053
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 634)
Brandenberger R., Wei, H., Zhang, S., Lei, S., Muraige, J., Fisk, G.J.,
Li, Y., Xu, C., Pang, R., Guegler, K., Rao, M.S., Mandalam, R.,
Labkowski, J. and Stanton, L.W.

TITLE

Transcriptional characterization elucidates signaling networks that
control human ES cell growth and differentiation

COMMENT

Nat. Biotechnol. 22 (6), 707-716 (2004)
Contact: Brandenberger R
Regenerative Medicine
Geron Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760

Email: rbrandenberger@geron.com
Insert length: 634 Std Error: 0.00.
Location/Qualifiers

FEATURES

source

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="embryonic stem cell, retinoic acid and
mitogen-treated hES cell line H7"
/clone_id="GRN_PRENBU"
/note="Oligo dT primed, full-length enriched cDNA library
from hES cell line H7 (p29) maintained in feeder-free
conditions. Embryoid bodies were generated in the presence
of all-trans retinoic and mitogens."

ORIGIN

Query Match 31.9%; Score 593; DB 7; Length 634;
Best Local Similarity 100.0%; Pred. No. 2,9e-270;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 66 AGTCGGGCGCGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 125
DB 42 AGTCGGGCGCGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 101
QY 126 CTGGGGGCTGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 185
DB 102 CTGGGGGCTGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 161
QY 186 CATGGCTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 245
DB 162 CATGGCTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 221
QY 246 CCGGCGCGCGTGTGCAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 305
DB 222 CCGGCGCGCGTGTGCAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 281
QY 306 CTTGGGCTCGAAGAGGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 365
DB 282 CTTGGGCTCGAAGAGGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 341
QY 366 CCGGAGACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 425
DB 342 CCGGAGACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 401
QY 426 GCGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485
DB 402 GCGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 461
QY 486 CCTGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 545
DB 462 CCTGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 521
QY 546 TAGAGCAACCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 605
DB 522 TAGAGCAACCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 581
QY 606 TGTCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 658
```

```
Db 582 TGTCCATGAGAGACAGCCGCGAGAGTGAACCTTCACTTGTATGACTTT 634
|||||
RESULT 14
LOCUS BI767278 627 bp mRNA linear EST 25-SEP-2001
DEFINITION 603057995F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5207614 5',
mRNA sequence.
ACCESSION BI767278 GI:15758869
VERSION BI767278.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 627)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Straubeberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM11520 row: m column: 23
High quality sequence stop: 580.
FEATURES
source
Location/Qualifiers
1..627
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5207614"
/lab_host="DH10B"
/clone_lib="NIH_MGC_122"
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source:
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleen. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."
ORIGIN
Query Match 31.3%; Score 582; DB 4; Length 627;
Best Local Similarity 100.0%; Pred. No. 4.9e-265;
Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 769 GGTAAAGCTCAACGTCGCCCCCGATGCGACCGACAGAGAGGCTCTTGTCAATCA 828
DB 1 GGTAAAGCTCAACGTCGCCCCCGATGCGACCGACAGAGAGGCTCTTGTCAATCA 60
QY 829 GGCTGACCTGCAAGAGCGGAGGCGCGAGAGAGACCAAGCCACTGAGGAGCTGGCGAG 888
DB 61 GGCTGACCTGCAAGAGCGGAGGCGCGAGAGAGACCAAGCCACTGAGGAGCTGGCGAG 120
QY 889 CTGGAGAGAGAGACGAGCGCGCTCAAGTTTCAAGGTTGACAGCGCGCTGGAGCAGTC 948
DB 121 CTGGAGAGAGAGACGAGCGCGCTCAAGTTTCAAGGTTGACAGCGCGCTGGAGCAGTC 180
QY 949 TGGCTGCTACCAACATTTGGGTAGATGAGAACATCGAGAGAGAAACCACTAATTAATCT 1008
DB 181 TGGCTGCTACCAACATTTGGGTAGATGAGAACATCGAGAGAGAAACCACTAATTAATCT 240
QY 1009 CCGCGGAGATAGAAAATACACGTCCTCAATTTGGGCTGGCTCCCTTCCGTGGCCAGAA 1068
|||||
```

```
Db 241 CGCGGGGATAGAAAACACTACACGTCCTCAATTTGGGCTGGCTCCCTTCCGTGGCCAGAA 300
QY 1069 GTGAAACTGCCCCCGGACCTTCATATCCACTCGATCTCGTCCCATAGCCGGAAGC 1128
DB 301 GTGAAACTGCCCCCGGACCTTCATATCCACTCGATCTCGTCCCATAGCCGGAAGC 360
QY 1129 CATCAACATCCCAACCGAAAGCCCAAGGCGGTGAGCCGGGCTCTTCCACTTCTTGA 1188
DB 361 CATCAACATCCCAACCGAAAGCCCAAGGCGGTGAGCCGGGCTCTTCCACTTCTTGA 420
QY 1189 CACCCCAATGCGCAAGGCTCTCAGAGCTCCAGCAACGAGCTCCGGGAGCAACGAGCGGAG 1248
DB 421 CACCCCAATGCGCAAGGCTCTCAGAGCTCCAGCAACGAGCTCCGGGAGCAACGAGCGGAG 480
QY 1249 CAAACACTTTGTGAGTCCCCCAAGGCCCAAGGCAAGAGTGTGGTGGGCAAGTGGC 1308
DB 481 CAAACACTTTGTGAGTCCCCCAAGGCCCAAGGCAAGAGTGTGGTGGGCAAGTGGC 540
QY 1309 CAGAGGGGCAAGAAACAGGCCCCCTCTGGGAGCCCGCATCC 1350
DB 541 CAGAGGGGCAAGAAACAGGCCCCCTCTGGGAGCCCGCATCC 582
|||||
RESULT 15
LOCUS BG104777 1027 bp mRNA linear EST 30-JAN-2001
DEFINITION 602311726F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4421335 5',
mRNA sequence.
ACCESSION BG104777
VERSION BG104777.1 GI:12598623
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1027)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Straubeberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM10160 row: d column: 08
High quality sequence stop: 692.
FEATURES
source
Location/Qualifiers
1..1027
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4421335"
/tissue_type="adrenal cortex carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_84"
/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site_1:
NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT
primed. Average insert size 1.225 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
ORIGIN
Query Match 30.9%; Score 574; DB 4; Length 1027;
Best Local Similarity 99.8%; Pred. No. 3e-261;
Matches 624; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 287 AGCTTGCCGCTGAGCGCTCTGCGCTCGAAGGCGATCGAGAGTGAATCGGAGACAG 346
DB 1 AGCTTGCCGCTGAGCGCTCTGCGCTCGAAGGCGATCGAGAGTGAATCGGAGACAG 60
|||||
```

OY	347	CGCTGCCGGGAGGTGTCTCGGAACTCCCGACAGCTGCGGTTGGCGGGCACATAGGCCGA	406
Db	61	CGTCCCGGGGGGTGTCTCGGGACCCCGCAGCTGCGGTTGGCGGGCACATAGGCCGA	120
OY	407	AGCACCCGGGAGCTGTGGGCGGCGGTGTGAGGACAGCTCAGCGGAGGAGGAGGAGC	466
Db	121	AGCACCCGGGAGCTGTGGGCGGCGGTGTGAGGACAGCTCAGCGGAGGAGGAGGAGC	180
OY	467	GACTTTCGGCTGGAGTGGCCCTGCTCTCTGAGAGACTGACGGGCTGGGACAGCGAGAT	526
Db	181	GACTTTCGGCTGGAGTGGCCCTGCTCTCTGAGAGACTGACGGGCTGGGACAGCGAGAT	240
OY	527	GAGAAAGANGTAGAGAGAGTGAAGCCGAACTTCGCGAGGCTTCAAGAGCAGCTGAAGTTT	586
Db	241	GAGAAAGANGTAGAGAGAGTGAAGCCGAACTTCGCGAGGCTTCAAGAGCAGCTGAAGTTT	300
OY	587	GAAAGGCTCCAGTGGGACGCTGTCCATGTAGGAGGAGACAGCGCGAGAGATGTGACCTTCACC	646
Db	301	GAAAGGCTCCAGTGGGACGCTGTCCATGTAGGAGGAGACAGCGCGAGAGATGTGACCTTCACC	360
OY	647	CTGTATGACTTTGACAACCAACGGCAAGTTCACCCGAGAGACATCAACAGCTTGTCTGCAC	706
Db	361	CTGTATGACTTTGACAACCAACGGCAAGTTCACCCGAGAGACATCAACAGCTTGTCTGCAC	420
OY	707	ACCATCTATAGAGTGGTGAATCTCTCTGTGCAACCACTCCGCCAACTTCAGCAAGATGCTG	766
Db	421	ACCATCTATAGAGTGGTGAATCTCTCTGTGCAACCACTCCGCCAACTTCAGCAAGATGCTG	480
OY	767	CGGGTAAAGCTCAGCGGTGGCCCCGATTTGGCAGCGAGAGAGGAGGCGTCTGTCAAT	826
Db	481	CGGGTAAAGCTCAGCGGTGGCCCCGATTTGGCAGCGAGAGAGGAGGCGTCTGTCAAT	540
OY	827	CAGGCTGACCTTGACAGAGCGCAAGGCCCCCAGAGAGAGACCAAGGCCCATGTAGGACCTTGCG	886
Db	541	CAGGCTGACCTTGACAGAGCGCAAGGCCCCCAGAGAGAGACCAAGGCCCATGTAGGACCTTGCG	600
OY	887	AGCTGGAGAGAGACGACGAGAGCCC	911
Db	601	AGCTGGAGAGAGACGACGAGAGCCC	625

Search completed: December 30, 2004, 03:20:00
Job time : 8798 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2004, 13:06:18 ; Search time 212 Seconds
(without alignments)
6232.814 Million cell updates/sec

Title: US-09-993-966-5

Perfect score: 1859
Sequence: 1 gaattcgccctctaatacgc.....aaaaaaaaaagcgcgcc 1859

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/prodata/1/ina/PCBUS.COMB.seq: *
6: /cgn2_6/prodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	1434.8	77.2	1438 4	US-09-506-066E-5 Sequence 5, Appl
2	1221.4	65.7	1731 4	US-09-506-066E-3 Sequence 3, Appl
3	252	13.6	1285 4	US-09-506-066E-7 Sequence 7, Appl
4	206.8	11.1	1307 4	US-09-506-066E-9 Sequence 9, Appl
5	57.6	3.1	7218 1	US-08-232-463-14 Sequence 14, Appl
6	53.8	2.9	3768 4	US-10-101-464A-949 Sequence 949, App
7	51.2	2.8	2790 4	US-09-254-594-5 Sequence 5, Appl
8	51.2	2.8	3432 4	US-09-254-594-4 Sequence 4, Appl
9	49.8	2.7	375 4	US-09-248-796A-7789 Sequence 7789, Ap
10	49.8	2.7	459 4	US-09-707-919A-13 Sequence 13, Appl
11	49.2	2.6	623 3	US-09-043-303-5 Sequence 5, Appl
12	48.4	2.6	5852 1	US-07-867-106-2 Sequence 2, Appl
13	48.4	2.6	4403765 3	US-09-103-840A-2 Sequence 2, Appl
14	48.2	2.6	516 4	US-09-083-268-1 Sequence 1, Appl
15	48.2	2.6	4481 4	US-09-041-886-18 Sequence 18, Appl
16	48.2	2.6	4481 4	US-09-648-281-1 Sequence 1, Appl
17	48.2	2.6	4481 4	US-09-707-919A-20 Sequence 20, Appl
18	48.2	2.6	4481 4	US-09-083-268-2 Sequence 2, Appl
19	47.6	2.6	3162 4	US-10-140-002-11 Sequence 11, App
20	47.4	2.5	4403765 3	US-09-103-840A-2 Sequence 2, Appl
21	47.4	2.5	4411529 3	US-09-103-840A-1 Sequence 1, Appl
22	47.2	2.5	1859 3	US-09-339-913-23 Sequence 23, Appl
23	47.2	2.5	1859 3	US-09-298-731-23 Sequence 23, Appl
24	47.2	2.5	1859 3	US-09-350-614-23 Sequence 23, Appl
25	47.2	2.5	1955 3	US-09-399-913-19 Sequence 19, Appl
26	47.2	2.5	1955 3	US-09-298-731-19 Sequence 19, Appl
27	47.2	2.5	1955 4	US-09-350-614-19 Sequence 19, Appl

28	47.2	2.5	2009 3	US-09-389-913-13 Sequence 13, Appl
29	47.2	2.5	2009 3	US-09-298-731-13 Sequence 13, Appl
30	47.2	2.5	2009 4	US-09-350-614-13 Sequence 16, Appl
31	46.6	2.5	430 4	US-09-621-976-16656 Sequence 1656, A
32	46.4	2.5	7055 4	US-09-976-594-241 Sequence 941, App
33	46	2.5	435 4	US-09-389-681-285 Sequence 285, App
34	46	2.5	435 4	US-09-620-405B-285 Sequence 285, App
35	46	2.5	435 4	US-09-339-338-285 Sequence 285, App
36	46	2.5	435 4	US-09-433-826B-285 Sequence 285, App
37	46	2.5	435 4	US-09-604-287A-285 Sequence 285, App
38	46	2.5	435 4	US-09-834-759-285 Sequence 285, App
39	46	2.5	435 4	US-09-590-751A-285 Sequence 285, App
40	46	2.5	1149 3	US-09-422-487-6 Sequence 6, Appl
41	46	2.5	1149 3	US-09-378-088A-128 Sequence 128, App
42	46	2.5	1149 4	US-09-643-596B-128 Sequence 128, App
43	46	2.5	1196 4	US-09-270-767-14804 Sequence 14804, A
44	45.6	2.5	152331 3	US-09-128-155-16 Sequence 16, Appl
45	45.4	2.4	528 4	US-09-232-785-242 Sequence 242, Appl

ALIGNMENTS

RESULT 1
US-09-506-066E-5
Sequence 5, Application US/09506066E
Patent No. 6630323
GENERAL INFORMATION:
APPLICANT: Scott, Matthew
APPLICANT: Wharton, Keith
APPLICANT: Zeng, Wenlin
TITLE OF INVENTION: Naked Cuticle Genes and their Uses
FILE REFERENCE: STAN-121
CURRENT APPLICATION NUMBER: US/09/506,066E
CURRENT FILING DATE: 2000-02-17
PRIOR APPLICATION NUMBER: 60/120,646
PRIOR FILING DATE: 1999-02-17
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 1438
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (8)...(1418)
OTHER INFORMATION: Nkdl coding sequence
US-09-506-066E-5

Query Match 77.2%, Score 1434.8, DB 4, Length 1438,
Best Local Similarity 99.9%, Pred. No. 0;
Matches 1436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	217	CCCAGCATGGGAACTTCACTCCAGCGCGCGCTGCAAGGAGGAGAGCC	276
DB	1	CCCAGCATGGGAACTTCACTCCAGCGCGCGCTGCAAGGAGGAGAGCC	60
QY	277	GGAAGTGAAGCTTCGCGCTGAGCGCTGCGGCTCGGAAGGGCATCGAGTGAAT	336
DB	61	GGAAGTGAAGCTTCGCGCTGAGCGCTGCGGCTCGGAAGGGCATCGAGTGAAT	120
QY	337	CGGAGACAGCGCTGCGCGCGGTGCTCGGACCCGACAGCTGCGGTGGCGGAC	396
DB	121	CGGAGACAGCGCTGCGCGCGGTGCTCGGACCCGACAGCTGCGGTGGCGGAC	180
QY	397	CATAGCCGAGAGCCCGGAGCTCGGGCGAGCTTGAAGACACGCTCAGCGAGA	456
DB	181	CATAGCCGAGAGCCCGGAGCTCGGGCGAGCTTGAAGACACGCTCAGCGAGA	240
QY	457	AGAGGAGAGCATTTGAGCTGAGAGTGCCTGCTCTGAGAGACTGA	516
DB	241	AGAGGAGAGCATTTGAGCTGAGAGTGCCTGCTCTGAGAGACTGA	300

QY	51	CAGCGAGACATAGAAAGAAATGAGAGAGTAAAGGAAACCCGAGCCAGGCTCCAAAGAAACA	576
Db	301	CAGCGAGATAGAAAGAAATGAGAGAGTAAAGCAACCCGAGCCCAAGAAACA	360
QY	577	GCTGAAGTTTAAAGAGCTCCAGTGGCAAGTGTCCATGAGAGAGGACAAACCGGACGAGAGTG	636
Db	361	GCTGAAGTTTAAAGAGCTCCAGTGGCAAGTGTCCATGAGAGAGGACAAACCGGACGAGAGTG	420
QY	637	GACCTTCAACCCCTGTATGACTTTGACAAACAGGCAAGGTCACCCGAGAGGACATCACAG	696
Db	421	GACCTTCAACCCCTGTATGACTTTGACAAACAGGCAAGGTCACCCGAGAGGACATCACAG	480
QY	697	CTTGCTGACACATCTATGAGAGGTGTGAGACTCTCTGTCAACCACTCCCAATCCAG	756
Db	481	CTTGCTGACACATCTATGAGAGGTGTGAGACTCTCTGTCAACCACTCCCAATCCAG	540
QY	757	CAAGATGCTGGGGGTTAAAGTCACCGTAGCCCCGATGGACCCAGACCAAGAGAGCGT	816
Db	541	CAAGATGCTGGGGGTTAAAGTCACCGTAGCCCCGATGGACCCAGACCAAGAGAGCGT	600
QY	817	CTTGTCAATCAAGCTGACTTGACAGAGCGCAAGGCCCGGACAGAGACCAAGCCCACTGA	876
Db	601	CTTGTCAATCAAGCGCTGACTTGACAGAGCGCAAGGCCCGGACAGAGACCAAGCCCACTGA	660
QY	877	GGAACCTGGGGAGCTGGGAGAAAGACAGACGAGCCCGCTCAGGTTCCAGGGTGAACGCG	936
Db	661	GGAACCTGGGGAGCTGGGAGAAAGACAGACGAGCCCGCTCAGGTTCCAGGGTGAACGCG	720
QY	937	CCTGAGACAGTCTGGCTCTACACCATTTGGTGTATAGAAACATCGAGAGAGAAACA	996
Db	721	CCTGAGACAGTCTGGCTCTACACCATTTGGTGTATAGAAACATCGAGAGAGAAACA	780
QY	997	CTACTTAGATCTCGCGGGGATAGAAACTACAGTCCCAATTTGGGCTGGCTCCCTTC	1056
Db	781	CTACTTAGATCTCGCGGGGATAGAAACTACAGTCCCAATTTGGGCTGGCTCCCTTC	840
QY	1057	CGTGGCCCAAGAGTGAAGACTGCCCCCGGACCTCCCAATCCCATCTCGTCTCCCA	1116
Db	841	CGTGGCCCAAGAGTGAAGACTGCCCCCGGACCTCCCAATCCCATCTCGTCTCCCA	900
QY	1117	TGAGCGGAAACCATCCATCCCAACCGAAAGGCCCAAGAGCGTGAACCGGACTCCCT	1176
Db	901	TGAGCGGAAACCATCCATCCCAACCGAAAGGCCCAAGAGCGTGAACCGGACTCCCT	960
QY	1177	CCACTTCTTTGACACCCCAATCGCCAAAGGTCAGAGCTCCAGCAACGAGCTCCGGGAGAC	1236
Db	961	CCACTTCTTTGACACCCCAATCGCCAAAGGTCAGAGCTCCAGCAACGAGCTCCGGGAGAC	1020
QY	1237	CCAGGACGGGAGACAGCACTTTGTGAGTCCCCCAAGGCCCAAGGCAAGATGTGTGGTGT	1296
Db	1021	TCAGGACGGGAGACAGCACTTTGTGAGTCCCCCAAGGCCCAAGGCAAGATGTGTGGTGT	1080
QY	1297	GGGCGACGTGGCGAGAGGGGGCAAGAAACAAGGCCCTGTGGGACCCCGCATGCCGTGGGT	1356
Db	1081	GGGCGACGTGGCGAGAGGGGGCAAGAAACAAGGCCCTGTGGGACCCCGCATGCCGTGGGT	1140
QY	1357	GTCCCCCTCGGCCCACTGCTGGCCAGGCCCGGACTCTCTCCCTTCCCTTAGCCCCCTTCG	1416
Db	1141	GTCCCCCTCGGCCCACTGCTGGCCAGGCCCGGACTCTCTCCCTTCCCTTAGCCCCCTTCG	1200
QY	1417	GCACAAAGACCAAGCACCCAGGCCCAAGAGAGCCAGAGGCTGCGGGGCTTGACAGGC	1476
Db	1201	GCACAAAGACCAAGCACCCAGGCCCAAGAGAGCCAGAGGCTGCGGGGCTTGACAGGC	1260
QY	1477	ACCAATGGAGCTCAGAGTGGCCCTGTCTGTGGGGCGGAGACCTGCGGGAGCTGCGCGCTT	1536
Db	1261	ACCAATGGAGCTCAGAGTGGCCCTGTCTGTGGGGCGGAGACCTGCGGGAGCTGCGCGCTT	1320
QY	1537	GGTGGTGTATGAGAGCCAGAGCCGGGACCCGGTCCAGAGACATGAGACACACACCA	1596
Db	1321	GGTGGTGTATGAGAGCCAGAGCCGGGACCCGGTCCAGAGACATGAGACACACACCA	1380
QY	1597	TGAACATCACCATTTACACCACTTTCACGACATAGAGCCCTTCCAGGGGCC	1654

[illegible]

QY 682 AGAGCATGACGAGCTTGTCTGACACCATCTATGAGGTGGACTCTCTCTGCAACCA 741
 Db 598 TGAGGACATTACAGAGCTTGTCTGATACCATCTATGAAAGGTGGTCTCTCTGGAACCA 657
 QY 742 CTCCTCCAAATCAGCAGCAAGATGCTGGGGTAAAGCTACCGTGGCCCCCGATGGCAGCA 801
 Db 658 TTCTCCCAATCAAGCAAGACACTGGGGGTGAAGCTACCGTGGCTCTGACGGAGCCA 717
 QY 802 GAGCAAGAGAGAGCGTCTCTGTCTCATCTAGGCTGACCTGCAAGCGCAAGCGCCGAGCA 861
 Db 718 GATTAAGAGAGAGCGTCTCTTTTCAACATACCATCTGCAAGGCAAGAGCGCCGAGCAGA 777
 QY 862 GACCAAGCCCACTGAGAGACTGCGAGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 921
 Db 778 CACCAACCCGCTGAGAGAGCTGCGTGGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 837
 QY 922 CAGAGGTGACAGCGCGCTGGAGAGAGTGTGGCTGTACACCATTTGGCTAGATGAGAACAT 981
 Db 838 CAGAGGTGACAGCGCGCTGGAGAGAGTGTGGCTGTACACCATTTGGCTAGATGAGAACAT 897
 QY 982 CGAG 1041
 Db 898 TGAG 957
 QY 1042 GCGTGGCT 1101
 Db 958 ACCGGAATCCCT 1017
 QY 1102 TCGATCTCGCT 1161
 Db 1018 TCGCT 1077
 QY 1162 GGAAG 1221
 Db 1078 GGAAG 1137
 QY 1222 AGCGGCT 1281
 Db 1138 AGCGGCT 1197
 QY 1282 CAG 1341
 Db 1198 CAG 1257
 QY 1342 CGCCATCT 1401
 Db 1258 CACCAATCT 1317
 QY 1402 CTTAGGCT 1461
 Db 1318 CTTAGGCT 1377
 QY 1462 CGGAG 1518
 Db 1378 CGGAG 1437
 QY 1519 GCGGAG 1578
 Db 1438 GAGGAG 1497
 QY 1579 TGAAG 1638
 Db 1498 CGAAG 1557
 QY 1639 CGCCT 1698
 Db 1558 CGCCT 1603
 QY 1699 TATTAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1758
 Db 1604 TATTAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1663

QY 1759 CACTAATATTAGTACGCTACATGATAGAGATCTATGAAAGACAGAACTAACTTTTA 1818
 Db 1664 CACTAATATTAGTACGCTACATGATAGAGATCTATGAAAGACAGAACTAACTTTTA 1723
 QY 1819 T 1819
 Db 1724 T 1724
 RESULT 3
 US-09-506-066E-7
 ; Sequence 7, Application US/09506066E
 ; Patent No. 6610323
 ; GENERAL INFORMATION:
 ; APPLICANT: Scott, Matthew
 ; APPLICANT: Wharton, Keith
 ; APPLICANT: Zeng, Wenlin
 ; TITLE OF INVENTION: Naked Cuticle Genes and their Uses
 ; FILE REFERENCE: STAN-121
 ; CURRENT APPLICATION NUMBER: US/09/506,066E
 ; CURRENT FILING DATE: 2000-02-17
 ; PRIOR APPLICATION NUMBER: 60/120,646
 ; PRIOR FILING DATE: 1999-02-17
 ; NUMBER OF SEQ. ID NOS: 15
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 7
 ; LENGTH: 1285
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (2)...(1282)
 ; OTHER INFORMATION: Nkd2 coding sequence
 US-09-506-066E-7
 Query Match 13.6%; Score 252; DB 4; Length 1285;
 Best Local Similarity 56.3%; Pred. No. 1.3e-47;
 Matches 631; Conservative 0; Mismatches 450; Indels 39; Gaps 7;
 QY 516 GCAAG 575
 Db 180 GCAAG 239
 QY 576 AGCTGAAGTTGAAGAGCTTCACTGCGACGTGTCTCATGAGAGAGAGAGAGAGAGAG 635
 Db 240 ACTTGAACATTTGACGACCTGACGTGTGACCTCTCACTGAGAGAGAGAGAGAGAGAG 299
 QY 636 GGAAG 695
 Db 300 GGAAG 359
 QY 696 GCTTGTGACACCATCTATGAG 755
 Db 360 GCTTGTGACACCATCTATGAG 419
 QY 756 GCAAG 815
 Db 420 GCAAG 476
 QY 816 TCTTGTGATCAAGGCTGACCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 875
 Db 477 GTCTTCTACTGAGCAAG 536
 QY 876 AGAAG 935
 Db 537 ATGAG 596
 QY 936 GCTTGAAG 995
 Db 597 CCGAG 656
 QY 996 ACTAATTAGATCTGCGCGGAGATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1055

Dh 657 ACTACCTAGACCTTGTGGATGAGAACTACATCTAAGTTTGTCTGTGGTCAACAC 716
Qy 1056 CCGTGGCCCAAGAGTCAAGATGCCCCCGGCACTC---CAATCCCATCTGATCTGCT 1112
Db 717 CTGAGCAGGCGCAAGCAAGAACTATATGCAAGGCGCAACATTTCCAAAGAGTCCGAT 776
Qy 1113 CCATGAGCGGAAAGCCATTCACATCCCAACCGAAAGCCCAAGGGGTGAGCCGGGCT 1112
Db 777 CACAGAGTGGATGCTCCACGCTTATACCAACGCGAGTCTCAAGTCTGG----- 827
Qy 1173 CTTTCCACTTCTTGTACACCCCAATCCCAAGGTCTGAGCTTCAGCAACGCTCCGG 1232
Db 828 --CTGACATGTCTATACAGCTATAGCTGCTGACCCGGGCTGTGCTGACAGCCCG 885
Qy 1233 GCACCCAGAGAGGAGCAAGCATTTGTAGGTCCCCCAGGCCCCCAGGCAAGGTGG 1292
Db 886 GATTAAGGGGCAAGAG--AAGCAGTCTCAAGTCTCTTAAGGGTCCAGAAACCTT-- 942
Qy 1293 GTGTGGGCAAGTGGCCAGAGGGGCAAGAAACAGAGCCCTCTGGACCCGCAATCCCTG 1352
Db 943 ---TGGACACACAGGCAAGTGGCAAGCAAGGAAAGCTTCAAGTATTTGCTGAGGCT 999
Qy 1353 CGGTGTCCCTCCGCCACCTGCTGCAAGCCCGGCTCTCTCCCTCTAGCCCCC 1412
Db 1000 GCCATTGCCCCAGAGTGTCTGAGATGGCAACCATTTCTGACCCCCACCAACAGCTCC 1059
Qy 1413 TCGGGCAAGAGCAAGCAAGCAAGCAAGAGAGCCAGAGAGGCTGCCGGGGCTGC 1472
Db 1060 ACCGAGCCCTATGTGTCA-----CAAGCGTACCGGCAAGAAAGCCAGAAAGGCC 1109
Qy 1473 AGGCAACACTGAGCTCTGAGTGGC-----CTGTCTGTGGGCGGAGCACTCGCGGAGC 1526
Db 1110 ACTTACCACTTAAGGGGCAAGGCAAGCTTACATGTGTGAGCATGAAAGTATGAGGAGC 1169
Qy 1527 TGCCCGCTTGTGTGTATGAGCAAGCGCGGCAAGCGGTCCAGAGCATGAGCAAC 1586
Db 1170 TGCTTCCATGTGTGGGCTGAGGGCTATGTGATGCTGTGTCTCAAGGGCATGAAACCC 1229
Qy 1587 ACCACCACTGAATCATCATCACTTACCACTTCTA 1626
Db 1230 ACCATCACTAGACACCAACCATCATCAACCA 1269

RESULT 4
US-09-506-066E-9
; Sequence 9, Application US/09506066E
; Patent No. 6630323
; GENERAL INFORMATION:
; APPLICANT: Scott, Matthew
; APPLICANT: Wharton, Keith
; APPLICANT: Zeng, Wenlin
; TITLE OF INVENTION: Naked Cuticle Genes and their Uses
; FILE REFERENCE: STAN-121
; CURRENT APPLICATION NUMBER: US/09/506,066E
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/120,646
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1307
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-506-066E-9

Query Match 11.1%, Score 206.8; DB 4; Length 1307;
Best Local Similarity 55.1%, Pred. No. 2.2e-37;
Matches 553; Conservative 0; Mismatches 412; Indels 39; Gaps 6;

Qy 670 CAAGGTCAACCCGAGAGGATCATCACAGCTTGTGACACCATCTATAGGTGTGAGCTC 729
Db 51 CGAGTCAACAGGAGGATGTCAAGCTCATGACACCATCTATAGGTGTGAGTGC 110

Qy 730 CTCTGCAACACACTCCCAATCATTCAGCAAGATGTCTGGGTAAAGTCAACGCTGAGCCC 789
Db 111 CTGGTCAACACTCTCTCGGAGAGAGCAAGACCTCTCGTGAAGTAAACCTGAGCCC 170
Qy 790 CGATGGAGCAAGAGCAAGAGAGCGTCTTGTCAATCAGCTGACCTGAGAGCGCAAG 849
Db 171 TGAGCCCTCCAGAAAGAGGAGGAGGCTCTCTGTGAGCAAGACCGGAGGCCACCCG 230
Qy 850 GCCCCAGAGAGACCAAGCCCACTGAGGACTGCGGAGCTGGGAGAGAGAGAGCGAGC 909
Db 231 TTGCAAGATGAGGAGTGAATGAGCAAGAGGCCAAGGGTGGTGAACAGAGGTTCTGC 290
Qy 910 CCGGCTCAGGTTCCAGGGTGAACAGCGCTGAGAGCAAGTGTGAGTCAACCAATTCGCT 969
Db 291 ACAAGTGAAGAGCCCAAGTACTGACCCCAAGCCCTGTCTGGAAGCGGGGGCTCTAGCT 350
Qy 970 AGATGAGAAATGAGAGAGAGAAACCACTTATGATCTGCGGAGATGAGAACTACAC 1029
Db 351 GAGAGAGAAACAGGAGAGGAGAAACCACTGAGCTGAGCTGCGCGGAGTGAACCTACAC 410
Qy 1030 GTCCCAATTTGGGCTTGGCTCCCTTCCGTGGCCAGAACTCAAGACTGCCCCCGCAC 1089
Db 411 GTCCAGATTGGCTTGGGCTTGGCTTCTCTGTGCAAGCAAGAGAGCCCAAGGCAAGGC 470
Qy 1090 CTCCAATCCC---ACTGATCTGCTCCCATGAGCCGGAAGCCATCCATCCACACCG 1146
Db 471 CTGGCACTTCAAGGCGCGGTCTCGCTCCAGAGAGCCAGATACATGCTTACACACCG 530
Qy 1147 AAAGCCCAAGAGGCTGAGCCCGGCTCTCTTCACTTCTTGAACCCCAATGCGCAAGT 1206
Db 531 CAGGTCAAGATGCTGGTGAACAGCTGCGAGCCCTGCAAGCTGCG---TGCCCGGAGC 587
Qy 1207 CTGAGACTCAAGCAAGCTCCGGGGCAACCGAGAGAGAGCAAGCACTTTGAGGTC 1266
Db 588 CTGGAACAGCAGCCCGCGGAAGGCGGAG-----AAGCAGTCTCAAGTC 638
Qy 1267 CCCCAGAGCCCAAGAGAGTGTGGTGTGAGGCAAGTGGCCAGAGAGGCGCAAGAAACA 1326
Db 639 CCCCAGAGGCTCCGGAGAGCGGCTGGGGTGCAGCCAGAGCAAGTCCGGAGAAAGCTT 698
Qy 1327 GCCCCTCTGAGAACCGGCATCTCGGCTGCTCCCTCTCGGCCATCTGCTGCAAGCCC 1386
Db 699 CAGCTACTACTGCTCGGCGCTGC-----TGCCGCCCAAGGCCCTCAAGAGCGGCACCA 752
Qy 1387 GGCCTCTCTCCCTCTCTAGCCCCCTCGGGGCAAGAAAGCAAGACCGAGCGCAAGGA 1446
Db 753 CTTCGCGAGCCCCCAACCGCACTTACGGGCAAGCGGTACCGCAAAAGGCAAGGA 812
Qy 1447 GAGCCAGAGAGGCTGCGGGGCTTGCAGGCAAGCACTGAGTGGCTGTCTGTGG 1506
Db 813 GGGCACTGCGCACTCAAGGCCCAACAGCTCAGCTGCGCAAG-----TGA 860
Qy 1507 GCGGAGCACTTGGAGCTGCTCCGCTTGTGTGTATGAGAGCCAGGCGGAGGCC 1566
Db 861 GCAAGAGGTGTGCGGAGCTGCGGCCCAAGCGAGAGAGGGCTCAACGGTGTGCAAGT 920
Qy 1567 GGTTCAGAGCATGAGAGCAACCAACCACTGAACATCACTACCACTT-----ACCA 1620
Db 921 GATTCAGAGGCAAGAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 980
Qy 1621 CTTCTACAGACATGAGAGCCCTCCCAAGGAGGCCCACTGCA 1664
Db 981 CTTTCAACCGTCTTGAAGGCCACTGCAAGCAACTGTGCTCCA 1024

RESULT 5
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINER, F.
; APPLICANT: FALKNER, F. G.


```

; TITLE OF INVENTION: gene variants
; FILE REFERENCE: US 443
; CURRENT APPLICATION NUMBER: US/09/707,919A
; CURRENT FILING DATE: 2003-02-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 459
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Description of
; OTHER INFORMATION: Artificial derived nucleotide sequence of allelic
; OTHER INFORMATION: variant of SCA2 gene
US-09-707-919A-13

```

```

Query Match      2.7%; Score 49.8; DB 4; Length 459;
Best Local Similarity 52.0%; Pred. No. 0.051;
Matches 146; Conservative 0; Mismatches 122; Indels 13; Gaps 1;

```

```

QY 1336 GGGAGCCGCGCATCGCTGCGGTGTCCTCCCTCGCCACCTGCTGCGCCGCGCTCTCT 1395
DB 83 GGGCGCTCCCGCCCTTGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 142
QY 1396 CCCCTCCCTAGCCCTCGGGGCAAGAGCAAGACAGCAGCAGCAGCAGCAGCAGCAGCA 1455
DB 143 GCGCGCGCGCAACCGCGCGC-----CTCTCCGCTCGCGCGCGCGCGCGCTCTCTCT 189
QY 1456 GGGCTGCGGGGCGCTGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1515
DB 190 GCGTTCGCGGCGTCTCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 249
QY 1516 CCGCGCGAGCTCGCGCGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1575
DB 250 GGGTATGGGCGCCCTCAGCATGTCTGTGAAGCCCGCAGCAGCAGCAGCAGCAGCAGCA 309
QY 1576 ACATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1616
DB 310 ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 350

```

```

RESULT 11
US-09-043-303-5
; Sequence 5, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; APPLICANT: SANPEI, Kazuhiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; FILE REFERENCE: Primers Therefor
; CURRENT APPLICATION NUMBER: US/09/043,303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 623
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (341)..(583)
; FEATURE:
; OTHER INFORMATION: Tsp-2
US-09-043-303-5

```

```

Query Match      2.6%; Score 49.2; DB 3; Length 623;
Best Local Similarity 51.8%; Pred. No. 0.08;
Matches 143; Conservative 1; Mismatches 119; Indels 13; Gaps 1;
QY 1336 GGGAGCCGCGCATCGCTGCGGTGTCCTCCCTCGCCACCTGCTGCGCCGCGCGCTCTCT 1395

```

```

DB 155 GGGCTTCCCGCCCTTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 214
QY 1396 CCCCTCCCTAGCCCTCGGGGCAAGAGCAAGACAGCAGCAGCAGCAGCAGCAGCAGCA 1455
DB 215 GCGCGCGCAACCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 261
QY 1456 GGGCTGCGGGGCGCTGAGGACACCACTGCGCTCAGGTGCGCTGTCTGTGGGCGGAGCA 1515
DB 262 GCGTTCGCGGCGTCTCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 321
QY 1516 CCGCGCGAGCTGCGCGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1575
DB 322 GGTGTATGGCCCTTACCATGTCTGTGAAGCCCGCAGCAGCAGCAGCAGCAGCAGCA 381
QY 1576 ACATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1611
DB 382 ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 417

```

```

RESULT 12
US-07-867-106-2/c
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526r18
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920625
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ 7187
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Feeney, Joanne Longo
; REGISTRATION NUMBER: 35,134
; REFERENCE/DOCKET NUMBER: RICE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; US-07-867-106-2

```

```

Query Match      2.6%; Score 48.4; DB 1; Length 5852;
Best Local Similarity 57.1%; Pred. No. 0.33;
Matches 88; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 1697 ATTATTTTCATTAAATTAATTTGTTATATGATGATTAATTTGTTAATTAATTAATTTGTTAC 1756
Db 5753 ATTGTATTTTTTTTATATATATGTTATTTGTTGTTGTTTTTTTACTATATTTTCATATTTT 5694

QY 1757 TCCACTAATATTAGCTAGCCTACATGTAGAGATCTATGGAAACAGAACTAACTTT 1816
Db 5693 TATTTAATAAATTATATTAATTTAAATTTAAATTAATAAAAAAAAAAAAAAAAAAATTT 5634

QY 1817 TATTTATATGTTAAAAAAAAAAAAAAAAAAAAA 1850
Db 5633 AAATTTAAAAAAAAAAAAAAAAAAAAA 5600

RESULT 13
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      2.6%; Score 48.4; DB 3; Length 4403765;
Best Local Similarity 50.0%; Pred. No. 6.1;
Matches 121; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 70 GGGCCGGCGCGACGGCGGCAGAAACCGCTCCCGGCGCCGCTCGGGGCTCCGCTCGGCTCG 129
Db 1637226 GGGCCGGCGGTGCCGTGTACCGGCGGTGCGGTGGGGTCCGCGGGTCCGCGGGCGG 1637167

QY 130 GGGGCTGCTTGGGAGGAGAGAGACCAAGGAGGCGCCAGGCGCGCGGCGCGGCGCATG 189
Db 1637166 CGGCTGATCACTGCTGCTGGGCGGCAAGGCGCGGCGCGGCGCGGCGGAGACCGGCGCGC 1637107

QY 190 GCTTAGGAGCGCTCCCGGCGCGCAGCCAGCCAGCATGGGAAACTTCACTTCAAGCCGCGC 249
Db 1637106 TGTGTGGCGTCCGGCGGTGACCGCGGAGCGGCGCCAAAGGCAACAGGCGCTTCAACGAGG 1637047

QY 250 CGCGCTGTGCAAGCGCAGGAGAGACCCGGAAGGTGACAGCTTCCGCTGAGCGCTGCGCTG 309
Db 1637046 TGCCGGCGCGTACCGGCGGCAACCGCGGCAACCCCGGACCGCGCGGCGCGGAGCAACCGG 1636987

QY 310 GG 311
Db 1636986 CG 1636985

RESULT 14
US-09-083-268-1
; Sequence 1, Application US/09083268
; Patent No. 6673535
; GENERAL INFORMATION:

```

APPLICANT: Pulst, Stefan M
TITLE OF INVENTION: NUCLEIC ACID ENCODING SPINOCEREBELLAR
TITLE OF INVENTION: ATAXIA-2 AND PRODUCTS RELATED THERETO
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Mueeling, Raasch & Gebhardt, P.A.
STREET: 119 No. 6673535th Fourth Street
CITY: Minneapolis
STATE: Minnesota
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent'n Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,268
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/727,084
FILING DATE: 08-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: McCormack, Myra H
REGISTRATION NUMBER: 36,602
REFERENCE/DOCKET NUMBER: 232.00010101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612/305-1220
TELEFAX: 612/305-1228
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 516 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: DNA (genomic)
US-09-083-268-1

[illegible]

```

NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSER: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 7000
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 2626
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 4481 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 163..4099
US-09-041-886-18

```

THIS PAGE IS BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2004, 13:06:18 ; Search time 12708 Seconds
(without alignments)
6917.812 Million cell updates/sec

Title: US-09-993-966-5

Perfect score: 1859
Sequence: 1 gaattcgccctctaactg.....aaaaaaaaaagcgcgcgc 1859

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBankl : *
1: gb_ba : *
2: gb_ncg : *
3: gb_in : *
4: gb_om : *
5: gb_ov : *
6: gb_pat : *
7: gb_pl : *
8: gb_pl : *
9: gb_pr : *
10: gb_ro : *
11: gb_sta : *
12: gb_sy : *
13: gb_un : *
14: gb_vl : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1786	96.1	1786	9	AY061883 Homo sapi
2	1763	94.8	2604	9	BC051288 Homo sapi
3	1610	86.6	1788	9	AB062886 Homo sapi
4	1438	77.4	1438	9	AF358135 Homo sapi
5	1434.8	77.2	1438	6	AR405917 Sequence
6	1259.6	67.8	1765	10	BC034838 Mus muscu
7	1221.4	65.7	1731	6	AR405916 Sequence
8	1194.8	64.3	1671	10	AF358134 Mus muscu
9	1078.4	58.0	1416	10	AF343352 Mus muscu
10	1066.6	57.4	1401	6	AX172351 Sequence
11	849.2	45.7	1053	6	CO721681 Sequence
12	783	42.1	16319	9	HSA303140 Homo sapi
13	783	42.1	16973	9	AC007608 Homo sapi
14	783	42.1	16973	9	AC007608 Homo sapi
15	775	41.7	17879	2	AC145238 Homo sapi
16	754.2	40.6	194869	2	AC145003 Pan trogl
17	754.2	40.6	194869	2	AC145003 Pan trogl
18	754.2	40.6	194869	2	AC145003 Pan trogl
19	637.8	34.3	139145	2	AC146326 Felis cat

20	634.8	34.1	199277	2	AC145445 Canis fam
21	602	32.4	181401	2	AC146663 Sus scrofa
22	592	31.8	208061	2	AC145254 Bos tauru
23	560.4	30.1	232509	2	AC134093 Rattus no
24	558.8	30.1	240965	2	AC126054 Mus muscu
25	272.4	14.7	1682	9	AF358137 Homo sapi
26	272.4	14.7	1798	9	AB062887 Homo sapi
27	255	13.7	1967	10	BC019952 Mus muscu
28	254.4	13.7	1947	10	AF358136 Mus muscu
29	252	13.6	1285	6	AR405918 Sequence
30	245.4	13.2	1885	6	AX780854 Sequence
31	232.8	12.5	1309	9	BC004940 Homo sapi
32	207	11.1	1358	9	BC012176 Homo sapi
33	206.8	11.1	1307	6	AR405919 Sequence
34	195.8	10.5	212531	2	AC145178 Gallus ga
35	169.2	9.1	1191	6	CO715890 Sequence
36	137	7.4	1950	6	AX557538 Sequence
37	136.6	7.3	2340	6	CO741468 Sequence
38	101.8	5.5	237332	2	AC131846 Rattus no
39	85	4.6	125020	9	AF429315 Homo sapi
40	75	4.0	99395	2	AC010446 Homo sapi
41	75	4.0	187289	9	AC116351 Homo sapi
42	73.4	3.9	181105	5	BX664719 zebrafish
43	70.2	3.8	99395	2	AC010446 Homo sapi
44	70	3.8	125020	9	AF429315 Homo sapi
45	68.6	3.7	208878	2	AC141571 Rattus no

ALIGNMENTS

RESULT 1	AY061883	1786 bp	MRNA	linear	PRI 26-DEC-2001
LOCUS	AY061883				
DEFINITION	Homo sapiens naked protein (NKD)		mRNA, complete cds.		
ACCESSION	AY061883				
VERSION	AY061883.1	GI:17978536			
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 1786)				
AUTHORS	Yan, D., Wiesmann, M., Rohan, V., Chan, V., Jefferson, A. B., Guo, L., Sakamoto, D., Caotien, R. H., Fuller, J. H., Reinhard, C., Garcia, P. D., Randazzo, F. M., Escobedo, J., Fantl, W. J. and Williams, L. T.				
TITLE	Elevated expression of axin2 and nkhd mRNA provides evidence that Wnt/beta -catenin signaling is activated in human colon tumors				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (26), 14973-14978 (2001)				
PUBMED	11752446				
REFERENCE	2 (bases 1 to 1786)				
AUTHORS	Rohan, M., Chan, V., Yan, D., Fantl, W. J. and Williams, L. T.				
TITLE	Direct Submission				
JOURNAL	Submitted (05-NOV-2001) Cancer Biology, Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608, USA				
FEATURES	Location/Qualifiers				
source	1..1786				
gene	/organism="Homo sapiens"				
CDS	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	1..1786				
	/gene="NKD"				
	159..1571				
	/gene="NKD"				
	/note="HNKD"				
	/codon_start=1				
	/product="naked protein"				
	/protein_id="AL32374.1"				
	/db_xref="GI:17978537"				
	/translation="MGKLSKPAVACRRRESPEGSFAVSAWARKGIEWIGRORCPGGVSGROLAIGTRSTREIVGLRPTLSSEEDDRLEVALPEPKDGGSGDEKKMERVSPCKSKOLKFEELQCDVSMEDSRQEWFTPLYDDNNKQVTRBDITSLHTIYEVVDSVNHSPSTSKMLKYLTAADGSGSKSVLVNQADLOSARPRATKTKTE				

ORIGIN

DIRSWEKKORAPLRFOGSRLEOSGCVHCHVDNIERRNHYLDLAGIENTSOFQPS
 PSYAKSELPRPTSNPTFRSHBEAHIHTRKQCDPDSFRLDPIAKUSBLQOR
 LRATODSSKHFVSPKQKSVGVHGRARAKRNPDPPIPAVSPAHIAASPALEP
 SLAPLGHRKHAKESQOGCRGLQAPLASGVPVLEHRELPALVYESQAGPOVQ
 RHEHHHHHHHHHYHRYQT"

Query Match 96.1%; Score 1786; DB 9; Length 1786;
 Best Local Similarity 100.0%; Pred. No. 2.3e-313; Indels 0; Gaps 0;
 Matches 1786; Conservative 0; Mismatches 0;

QY 66 AGTCGAGCCCGGCGAGCGGCGAGAGCGGTCCTCCGCGCGCGCTCGGCTCCGCTCGG 125
 Db 1 AGTCGAGCCCGGCGAGCGGCGAGAGCGGTCCTCCGCGCGCGCTCGGCTCCGCTCGG 60
 QY 126 CTGCGGAGGCTGCTTCTCGGAGAGAGAGCCAGAGAGAGCGCGAGCGCGCGCGAG 185
 Db 61 CTGCGGAGGCTGCTTCTCGGAGAGAGAGCCAGAGAGAGCGCGAGCGCGCGAGCGG 120
 QY 186 CATGGCTTAGAGAGAGCTCCGCGCGCGCGAGCGCGCGAGAGAGAGAGAGAGAGAG 245
 Db 121 CATGGCTTAGAGAGAGCTCCGCGCGCGCGAGCGCGCGAGAGAGAGAGAGAGAGAG 180
 QY 246 CGGCGCGCGCTGTGCAAGCGAGAGAGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 305
 Db 181 CGGCGCGCGCTGTGCAAGCGAGAGAGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 240
 QY 306 CTTGGGCTCGAG 365
 Db 241 CTTGGGCTCGAG 300
 QY 366 CGGAGACCCGACAGCTGCGGTCGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 425
 Db 301 CGGAGACCCGACAGCTGCGGTCGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 QY 426 GCGAGCTGTGAG 485
 Db 361 GCGAGCTGTGAG 420
 QY 486 CCTGCTCTCTGAG 545
 Db 421 CCTGCTCTCTGAG 480
 QY 546 TGAAGCAACCTGCTCGAG 605
 Db 481 TGAAGCAACCTGCTCGAG 540
 QY 606 TGTTCATGAG 665
 Db 541 TGTTCATGAG 600
 QY 666 ACCGCAAGGTCACCCGAG 725
 Db 601 ACCGCAAGGTCACCCGAG 660
 QY 726 ACTGCTCTGTCAACCACTCCCAACATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 785
 Db 661 ACTGCTCTGTCAACCACTCCCAACATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
 QY 786 CCCCCTGAG 845
 Db 721 CCCCCTGAG 780
 QY 846 CAAGGCCCCGAG 905
 Db 781 CAAGGCCCCGAG 840
 QY 906 GAGGCCCCGCTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 965
 Db 841 GAGGCCCCGCTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 900
 QY 966 GCGTAGATGAG 1025

Db 901 GCGTAGATGAG 960
 QY 1026 ACACGTCCTCAATTTGGGCTGGCTCCCCCTTCCGTGGCCGAGAGTGAAGATGAGATGAG 1085
 Db 961 ACACGTCCTCAATTTGGGCTGGCTCCCCCTTCCGTGGCCGAGAGTGAAGATGAGATGAG 1020
 QY 1086 GCACCTCCCAATCCCACTCGATCGCTCCCATGAGCGGAGAGAGAGAGAGAGAGAGAGAG 1145
 Db 1021 GCACCTCCCAATCCCACTCGATCGCTCCCATGAGCGGAGAGAGAGAGAGAGAGAGAG 1080
 QY 1146 GAAAGCCCCAAGGCGTGGACCCGCGCTCTTCCACTTCTTGAACACCCCAATGCGCAAG 1205
 Db 1081 GAAAGCCCCAAGGCGTGGACCCGCGCTCTTCCACTTCTTGAACACCCCAATGCGCAAG 1140
 QY 1206 TCTCAGAGCTTCAGCAACCGCTCCGGGGGACCCAGAGACGGGAGACAGCACTTTGTAGGT 1265
 Db 1141 TCTCAGAGCTTCAGCAACCGCTCCGGGGGACCCAGAGACGGGAGACAGCACTTTGTAGGT 1200
 QY 1266 CCCCAGAGCCGAGGCAAGAGTGTGGGTGTGGGCGACGTCGCGCAGAGAGGGGCAAGAAACA 1325
 Db 1201 CCCCAGAGCCGAGGCAAGAGTGTGGGTGTGGGCGACGTCGCGCAGAGAGGGGCAAGAAACA 1260
 QY 1326 AGCCCCCTCTGGAGACCCGCGATCCCTGCGGTGTCCCTTCCGCTCCAGAGAGAGAGAG 1385
 Db 1261 AGCCCCCTCTGGAGACCCGCGATCCCTGCGGTGTCCCTTCCGCTCCAGAGAGAGAGAG 1320
 QY 1386 CGGCTCTCTCTCCCTCCCTAGCCCTCCGGGGGACAAAGAGAGAGAGAGAGAGAGAGAG 1445
 Db 1321 CGGCTCTCTCTCCCTCCCTAGCCCTCCGGGGGACAAAGAGAGAGAGAGAGAGAGAGAG 1380
 QY 1446 AGAGCAG 1505
 Db 1381 AGAGCAG 1440
 QY 1506 GCGGAGAGACCTGCGGAGAGCTGCGGCTCTTGTGTGTATGAGAGAGAGAGAGAGAG 1565
 Db 1441 GCGGAGAGACCTGCGGAGAGCTGCGGCTCTTGTGTGTATGAGAGAGAGAGAGAGAGAG 1500
 QY 1566 CGGTCCAG 1625
 Db 1501 CGGTCCAG 1560
 QY 1626 ACCAGACATAG 1685
 Db 1561 ACCAGACATAG 1620
 QY 1686 CACCACAGGCAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATA 1745
 Db 1621 CACCACAGGCAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATA 1680
 QY 1746 ATTATTGTTATCTCAGCAATATTAGTGTAGCTACAGTGAAGAGATCTATGAGAGAGAG 1805
 Db 1681 ATTATTGTTATCTCAGCAATATTAGTGTAGCTACAGTGAAGAGATCTATGAGAGAGAG 1740
 QY 1806 AACTAACTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATA 1851
 Db 1741 AACTAACTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATA 1786

RESULT 2
 BC051288 2604 bp mRNA linear PRI 30-JUN-2004
 LOCUS
 DEFINITION Homo sapiens naked cuticle homolog 1 (Drosophila), mRNA (cDNA clone
 WCC:5966 IMAGE:6646089), complete cds.
 ACCESSION BC051288
 VERSION BC051288.1 GI:30410965
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 2604)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,

REMARK COMMENT	JOURNAL PUBMED REFERENCE AUTHORS TITLE	TITLE
	Klausner R.D., Collins F.S., Wagner K.H., Shemmen C.M., Schuler G.D., Altschuler S.F., Zeeberg B., Buetow K.L., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heich F., Datchenko L., Marusina K., Farmer A.G., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T., Leung K., Scheetz T.E., Brownstein M.J., Uedini T.B., Toshitsuki S., Carninci P., Prange C., Raha S.S., Loquatlano N.A., Peters G.J., Abramson R.D., Mullinhy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Wolley K.C., Hale S.C., Garcia A.M., Gay L.J., Hillyk S.W., Villallon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shetchenko Y., Bonfield G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Gilmwood J., Schmitt J., Myers R.M., Butcherfield Y.S., Krzywiński M.I., Skalka U., Smalls D.E., Schercher A., Schein J.E., Jones S.J. and Marra M.A.	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
	2 (bases 1 to 2604) 12477932	
	Strausberg R. Direct Submission Submitted (25-APR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	
	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk	

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/IMG at: <http://image.llnl.gov>
Series: IRAL Plate: 47 Row: 0 Column: 23
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 20127660

```

FEATURES
source
Location/Qualifiers
1..2604
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:59666 IMAGE:6646089"
/tissue_type="Ovary, teratocarcinoma"
/clone_lib="NH_MGC_109"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
1..2604
/gene="NKD1"
/db_xref="locusID:85407"
/db_xref="MIM:607851"
225..1637
/gene="NKD1"
/codon_start=1
/product="naked cuticle homolog 1"
/protein_id="AAH51288.1"
/db_xref="GI:30410966"

```

```

/db_xref="IMM:607851"
/translation="MGKLSKPRPAVCCKRRSPEDSDPAVSAAMARKGIEEIGNORCE
GGVSGPQRLACTIGRSTRELVDGLDVLITSEEDBEFRLEVALPEKTDITLSDGB
KMERVSEPCGSKKQKPEELQCDVMEBDSROEMFTPLDPNNNGKVTEDTSLIL
HTTYEVDSDVNSPTSSKMLRVKLTVAPOGOSKRALVNOALQSLNTPAATPTPE
DLRSEKKORAPLARFGQSDLESGCCVHNCVDENIEBRNHYLDLAGIENTYSGPGS
PSVAQKSELPRPTSNPTFRSSHEBEALHITPRKQGVDPASFIHLDTPIAVSELQOR
LRIGQDSKHFVRSPKAOGSGVGVHARGARKNRPGRPAI PAVSPSHILASAPLLE
SLAPLHGKKHKRAKESQCGCRGLQAPLACGSPVLGRENRLPALVYESOAGCPVQ
RHEHHHHHHHHHHFYQT"

```

OY	66	AGTGGGCGCGGCGACGGCGCGCAGAGCGGTCGCGGCGCGCTCGGCGCTCGCTCGG	125
Db	67	AGTCGGGCGCGGCGACGGCGCGGAGCGGTCGCGGCGCGCTCGGCGCTCGCTCGG	126
OY	126	CTCGGGGCGTCTTCGGGAGGAGAGCCAGGGAGGCGCAGGCCCGGGGCGGGCG	185
Db	127	CTCGGGGCGTCTTCGGGAGGAGAGCCAGGGAGGCGCAGGCCCGGGGCGGGCG	186
OY	186	CATGCGTTAGGGAGCGCTCCCGGCGCGCGACGCCCAGCATGGGGAACTTCATCTCAAGC	245
Db	187	CATGCGTTAGGGAGCGCTCCCGGCGCGCGACGCCCAGCATGGGGAACTTCATCTCAAGC	246
OY	246	CGGCGCGCGTGTGCAAGCGCAGGGAGAGCCCGGAAAGTGACGCTTGCGGTGACGCTTG	305
Db	247	CGGCGCGCGTGTGCAAGCGCAGGGAGAGCCCGGAAAGTGACGCTTGCGGTGACGCTTG	306
OY	306	CCTGGGCTCGGAAAGGCGATCGAGGAGATGAGATCGGAGAGCAGCGCTGCGCCGGGCGTGTCT	365
Db	307	CCTGGGCTCGGAAAGGCGATCGAGGAGATGAGATCGGAGAGCAGCGCTGCGCCGGGCGTGTCT	366
OY	366	CGGGACCCCGACAGCTCGGTTGGCGGGCACCATAGGCGGAGCACCCGGGAGCTCGTGG	425
Db	367	CGGGACCCCGACAGCTCGGTTGGCGGGCACCATAGGCGGAGCACCCGGGAGCTCGTGG	426
OY	426	GGGACGTGTGAGAGACACGCTCGAGCGAGAGAGAGAGAGACATTCGCGCTGGAAAGTGG	485
Db	427	GGGACGTGTGTAGAGACACGCTCGAGCGAGAGAGAGAGAGACATTCGCGCTGGAAAGTGG	486
OY	486	CCCTGCTCTCTGAGAGAGATGACGGGCTGGGCGACCGGAGATGAGAGAGATGAGAGAG	545
Db	487	CCCTGCTCTCTGAGAGAGATGACGGGCTGGGCGACCGGAGATGAGAGAGATGAGAGAG	546
OY	546	TGAGCGAACCTGCTCCGAGGCTCCAAAGAGCAGCTGAAGTTGAAAGAGTCCAGTGGCAGC	605
Db	547	TGAGCGAACCTGCTCCGAGGCTCCAAAGAGCAGCTGAAGTTGAAAGAGTCCAGTGGCAGC	606
OY	606	TGTCCATGAGAGAGACAGCCGGCAGGAGTGAACCTTCACCCCTGTATGACTTTGACAACA	665
Db	607	TGTCCATGAGAGAGACAGCCGGCAGGAGTGAACCTTCACCCCTGTATGACTTTGACAACA	666
OY	666	ACGGCAAGGTCAACCGAGAGGACATCACAGCTTGCGACACCATCTATAGGTTGGTG	725
Db	667	ACGGCAAGGTCAACCGAGAGGACATCACAGCTTGCGACACCATCTATAGGTTGGTG	726
OY	726	ACTCTCTGTCAACCATCCCAATCCAGCAAGATGTGCGGGTAAAGTCAACCGTGG	785
Db	727	ACTCTCTGTCAACCATCCCAATCCAGCAAGATGTGCGGGTAAAGTCAACCGTGG	786
OY	786	CCCCCGATGCGAGCCAGACAAAGAGAGCGTCTTGTCAATCAGGCTGACCTGCGAGACG	845
Db	787	CCCCCGATGCGAGCCAGACAAAGAGAGCGTCTTGTCAATCAGGCTGACCTGCGAGACG	846
OY	846	CAAGGCCCGAGCAGACAAAGCCCATGAGAGCTGCGGAGGTGGGAGAGAAAGCAGC	905
Db	847	CAAGGCCCGAGCAGACAAAGCCCATGAGAGCTGCGGAGGTGGGAGAGAAAGCAGC	906

```

QY 906 GAGCGCCGCTCAGGTTCCAGGGTGACAGCCGCTGAGACAGTGTGCTGCTACCAACATT 965
DB 907 GAGCGCCGCTCAGGTTCCAGGGTGACAGCCGCTGAGACAGTGTGCTGCTACCAACATT 966
QY 966 GCGTAGATGAGAACATGAGAGAGAAACCACTAGATCTGCGCGGATAGAAAAT 1025
DB 967 GCGTAGATGAGAACATGAGAGAGAAACCACTAGATCTGCGCGGATAGAAAAT 1026
QY 1026 AACAGTCCCAATTTGGGCTGGGCTCCCTCCGCGGCCAAGATTCAGAACTGCCCCC 1085
DB 1027 AACAGTCCCAATTTGGGCTGGGCTCCCTCCGCGGCCAAGATTCAGAACTGCCCCC 1086
QY 1086 GCACTCTCAATCCCACTGATCTGCTCCCATGAGCGGAGAGCATCATCCACACC 1145
DB 1087 GCACTCTCAATCCCACTGATCTGCTCCCATGAGCGGAGAGCATCATCCACACC 1146
QY 1146 GAAAGCCCCAAGCGGTGAGACCGGCTCTCTTCACTTCTTTGACACCCCAATGCGCAAG 1205
DB 1147 GAAAGCCCCAAGCGGTGAGACCGGCTCTCTTCACTTCTTTGACACCCCAATGCGCAAG 1206
QY 1206 TCTCAGAGCTCCAGCAACGGCTCCGGGGACCCAGAGCGGAGAGACATTTGTGAGT 1265
DB 1207 TCTCAGAGCTCCAGCAACGGCTCCGGGGACCCAGAGCGGAGAGACATTTGTGAGT 1266
QY 1266 CCCCAGAGCCCAAGGAGAGTGTGGGTGTGGGCACTGTGGCCAGAGGGGCAAGAAACA 1325
DB 1267 CCCCAGAGCCCAAGGAGAGTGTGGGTGTGGGCACTGTGGCCAGAGGGGCAAGAAACA 1326
QY 1326 AGCCCCCTTGGGACCCGCGCATTCCTGTGGTCCCTCCGACCTGTGCTCCAGCC 1385
DB 1327 AGCCCCCTTGGGACCCGCGCATTCCTGTGGTCCCTCCGACCTGTGCTCCAGCC 1386
QY 1386 CCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1445
DB 1387 CCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1446
QY 1446 AGAGCCAGAGAGGCTGCGCGGGGCTGAGAGGACCACTGAGTGTGGCTGTCTCTG 1505
DB 1447 AGAGCCAGAGAGGCTGCGCGGGGCTGAGAGGACCACTGAGTGTGGCTGTCTCTG 1506
QY 1506 GCGCGGAGACCTTGCGGAGCTGCGCGCTTGTGTGTATGAGAGCCAGGCGGAGAC 1565
DB 1507 GCGCGGAGACCTTGCGGAGCTGCGCGCTTGTGTGTATGAGAGCCAGGCGGAGAC 1566
QY 1566 CCGTCCAGAGACATGAGACCAACCAACCATTAATTAATTAATTAATTAATTAATTA 1625
DB 1567 CCGTCCAGAGACATGAGACCAACCAACCATTAATTAATTAATTAATTAATTAATTA 1626
QY 1626 ACCAGACATAGAGCCCTCCCGAGGGCCCACTGCGCATGTGAAGAGCCCAACCCCGA 1685
DB 1627 ACCAGACATAGAGCCCTCCCGAGGGCCCACTGCGCATGTGAAGAGCCCAACCCCGA 1686
QY 1686 CACCAAGGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1745
DB 1687 CACCAAGGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1746
QY 1746 ATTATTTGTACTCCACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1805
DB 1747 ATTATTTGTACTCCACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1806
QY 1806 AACTAACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1828
DB 1807 AACTAACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1829

```

```

RESULT 3
LOCUS AB062886 1788 bp mRNA linear PRI 20-OCT-2001
DEFINITION Homo sapiens mRNA for Dvl-1-binding protein NKD1, complete cds.
ACCESSION AB062886
VERSION AB062886.1 GI:16303259
KEYWORDS
SOURCE Homo sapiens (human)

```

```

ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
TITLE Molecular cloning, gene structure, and expression analyses of NKD1
JOURNAL Int. J. Oncol. 19 (5), 963-969 (2001)
MEDLINE 21490203
PUBMED 11604995
REFERENCE 2 (bases 1 to 1788)
AUTHORS Katoh, M.
JOURNAL Direct Submission
TITLE Submitted (08-JUN-2001) Masaru Katoh, National Cancer Center
JOURNAL Research Institute, Genetics and Cell Biology Section; Tsukiji
5-chome, Chuo-ku, Tokyo 104-0045, Japan [E-mail: mkatoh@ncc.go.jp,
Tel:81-3-3542-2511]
FEATURES
source
1..1788
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="human fetal tissues"
1..1788
/gene="NKD1"
/codon_start=1
/product="Dvl-1-binding protein NKD1"
/protein_id="BAB70500.1"
/db_xref="GI:16303260"
/translat="MGKLSKPAVAVCKRRSPSPGSPAVSAWARKGIEITGRPC
GVSGRRLAETIGRSTRELIVDLRDLSEEDDPFLDLPALPEKTDGSGDE
KRAERVSEPCSGKKQKFEELQDYSMEDSDEMTFTLIDNNKATREDITSL
HTLYEVADSVVNSPTSSXKLRYKLTVAPDSKKSVLVNOADLOSARPAETKPE
DLRSWEKQAPLRFOGDSRLBOGCVHNCVDENIRRNHYLDLAGIENYSGFPGS
PSVAQSELPPTSPNPTRSRSHPEALIHPIHRKQCVDPASFFLDTPIAKVSELQOR
LRGTODGSKHFVSPRAOGKSVGVGVARGRKPKLGPALPAVSPSALHASPALLP
SLAPLHKHKHKAHAKESQOGRGLQAPLASGFLVLRHRLPALVYVESQAGPVPQ
RHEHHHHHHHHHHFYTQ"
ORIGIN
Query Match 86.6%; Score 1610; DB 9; Length 1788;
Best Local Similarity 100.0%; Pred. No. 16-281;
Matches 1610; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 219 CAGCATGAGGAAACTTCACTCAAGCGGCGCGCTGTGTCAAAGCGAGGAGAGCCCG 278
DB 1 CAGCATGAGGAAACTTCACTCAAGCGGCGCGCTGTGTCAAAGCGAGGAGAGCCCG 60
QY 279 AAGGTGACAGCTTCCCGTGAAGGCTGCTGGGCTCGGAAGGCAATGAGAGTGTG 338
DB 61 AAGGTGACAGCTTCCCGTGAAGGCTGCTGGGCTCGGAAGGCAATGAGAGTGTG 120
QY 339 GGAAGACGCGCTCCCGGCGGTGTCTCGGGAACCCGACAGCTGCGGTTGGCGGACCA 398
DB 121 GGAAGACGCGCTCCCGGCGGTGTCTCGGGAACCCGACAGCTGCGGTTGGCGGACCA 180
QY 121 GGAAGACGCGCTCCCGGCGGTGTCTCGGGAACCCGACAGCTGCGGTTGGCGGACCA 180
DB 399 TAGCGCGAAGCACCGGAGACTGTGGGCGACGTGTGAGAGACACGCTCAGCGAGAG 458
QY 181 TAGCGCGAAGCACCGGAGACTGTGGGCGACGTGTGAGAGACACGCTCAGCGAGAG 240
DB 459 AGAGAGACAGCTTCCGCTGGAAGTGCTGCTCTGTGAGAGAGCTGACGGGCTGGGCA 518
QY 241 AGAGAGACAGCTTCCGCTGGAAGTGCTGCTCTGTGAGAGAGCTGACGGGCTGGGCA 300
DB 519 GCGGAGATGAGAGAGATGAGAGATGAGAGAGCAACCTCCAGGCTCCAAGAGAGC 578
DB 301 GCGGAGATGAGAGAGATGAGAGATGAGAGAGCAACCTCCAGGCTCCAAGAGAGC 360
QY 579 TGAAGTTTGAAGAGCTTCAGTGCAGCTGTCCATGAGAGAGAGCAGCGGACAGAGTGA 638
DB 361 TGAAGTTTGAAGAGCTTCAGTGCAGCTGTCCATGAGAGAGAGCAGCGGACAGAGTGA 420

```

QY 639 CCTTCACTCTGTATGACTTTGACAAACGCGAAGTCAACCCGAGAGACATCAACGCT 698
 DB 421 CCTTCACTCTGTATGACTTTGACAAACGCGAAGTCAACCCGAGAGACATCAACGCT 480
 QY 699 TGTGACACACATCTATGAGTGTGTGACTCTCTGTCAACCAATCCCAACATCCGACA 758
 DB 481 TGTGACACACATCTATGAGTGTGTGACTCTCTGTCAACCAATCCCAACATCCGACA 540
 QY 759 AGATGCTGCGGTAAAGCTCAACCTGAGCCCGCATGAGCAGCAGAGAGAGAGCGCTCC 818
 DB 541 AGATGCTGCGGTAAAGCTCAACCTGAGCCCGCATGAGCAGCAGAGAGAGAGCGCTCC 600
 QY 819 TTGTCAATCAGGCTGACCTGAGAGCGCAAGGCCCGAGCAGAGACCAAGCCCATGAGG 878
 DB 601 TTGTCAATCAGGCTGACCTGAGAGCGCAAGGCCCGAGCAGAGACCAAGCCCATGAGG 660
 QY 879 ACCGCGGAGCTGGGAG 938
 DB 661 ACCGCGGAGCTGGGAG 720
 QY 939 TGGAGCAGTGTGCTGTACACCATTTGCTAGATGAGAAATCATGAGAGAGAGAAACCAT 998
 DB 721 TGGAGCAGTGTGCTGTACACCATTTGCTAGATGAGAAATCATGAGAGAGAGAAACCAT 780
 QY 999 ACTTAAATCTGCGCGGATGAGAAATCAACGTCCTCAATTTGGGCTGTGCTCCCTTCCG 1058
 DB 781 ACTTAAATCTGCGCGGATGAGAAATCAACGTCCTCAATTTGGGCTGTGCTCCCTTCCG 840
 QY 1059 TGGCGGAGAGTGAAGTGGCGCGCGCATCTCCATCTCCATCTCCATCTCCATCTCCAT 1118
 DB 841 TGGCGGAGAGTGAAGTGGCGCGCGCATCTCCATCTCCATCTCCATCTCCATCTCCAT 900
 QY 1119 AGCGGAGAGCATCAATCCCAACCGGAAAGCCCGAGAGAGAGAGAGAGAGAGAGAG 1178
 DB 901 AGCGGAGAGCATCAATCCCAACCGGAAAGCCCGAGAGAGAGAGAGAGAGAGAGAG 960
 QY 1179 ACTTCTTGAACACCCCAATGCGCAAGTCTCAAGATCTCAAGCAAGGCTCCGAGGCA 1238
 DB 961 ACTTCTTGAACACCCCAATGCGCAAGTCTCAAGATCTCAAGCAAGGCTCCGAGGCA 1020
 QY 1239 AGGAGCGGAG 1298
 DB 1021 AGGAGCGGAG 1080
 QY 1299 GCGCAGTGTGAG 1358
 DB 1081 GCGCAGTGTGAG 1140
 QY 1359 CCCCCTCCGCGCACTGTGCTGCAAGCCGCGCTCTCCCTCCCTGAGCCCTCCGAGG 1418
 DB 1141 CCCCCTCCGCGCACTGTGCTGCAAGCCGCGCTCTCCCTCCCTGAGCCCTCCGAGG 1200
 QY 1419 ACAAG 1478
 DB 1201 ACAAG 1260
 QY 1479 CACTGAGCTGAGTGTGCTGTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1538
 DB 1261 CACTGAGCTGAGTGTGCTGTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 QY 1539 TGGTGTATGAG 1598
 DB 1321 TGGTGTATGAG 1380
 QY 1599 AATATCAACCACTTACCACTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1658
 DB 1381 AATATCAACCACTTACCACTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
 QY 1659 CTGCAATATGAG 1718
 DB 1441 CTGCAATATGAG 1500

QY 1719 TTATATGATGATATATTTATTAATTAATTTATTTACTCTCACTAATATTTAGTACCT 1778
 DB 1501 TTATATGATGATATATTTATTAATTAATTTATTTACTCTCACTAATATTTAGTACCT 1560
 QY 1779 ACATGTGAGAGATCTATGAGAAACAGAGACTAACTTTATTTATTTATTTATTT 1828
 DB 1561 ACATGTGAGAGATCTATGAGAAACAGAGACTAACTTTATTTATTTATTTATTTAT 1610
 RESULT 4
 AF358135
 LOCUS
 DEFINITION
 Hemo sapiens naked cuticle-1 (NKD1) mRNA, complete cds.
 ACCESSION
 AF358135
 VERSION
 AF358135.1
 KEYWORDS
 GI:14211713
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE
 1 (bases 1 to 1438)
 Wharton, K.A., Jr., Zimmermann, G., Rousset, R., and Scott, M.P.
 Vertebrate proteins related to Drosophila Naked Cuticle bind
 Dishevelled and antagonize Wnt signaling
 Dev. Biol. 234 (1), 93-106 (2001)
 JOURNAL
 MEDLINE
 PUBMED
 11356022
 2 (bases 1 to 1438)
 Wharton, K.A., Jr., Zimmermann, G., and Scott, M.P.
 Direct Submission
 Submitted (08-MAR-2001) Pathology and Molecular Biology, University
 of Texas Southwestern Medical School, 5323 Harry Hines Blvd.,
 Dallas, TX 75390-9072, USA
 FEATURES
 source
 location/Qualifiers
 1..1438
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="16"
 /map="16q12"
 1..1438
 /gene="NKD1"
 8..1420
 /gene="NKD1"
 /note="contains Efx domain; similar to Drosophila
 melanogaster naked cuticle (nkd) and to Mus musculus Nkd1"
 /codon_start=1
 /product="naked cuticle-1"
 /protein_id="AAK57484.1"
 /db_xref="GI:14211714"
 /translation="MGKLSKPAVAVCKRRSPGDSFPAVSAAWARKGIEWIGRORCP
 GGVSGRDLRLAGTIGRSTRELVDVLRDTLSEEDDFELVALPPEKTDGAGSDE
 KKMVERSEPPGSKQLKFEELQCDVSMEDSDQEWPTLYDFDNNKVTREDITSIL
 HTLYEVLDSPSVNHSPTSSSKMLRYKLTVAARDGSKSVLVNADLDGASRRATKPRE
 DLSMKKKQAPRLRFQDGRDLDSGCTYHNCVDNTERKRYLTLAGIENTTSQPGPS
 PSVAQSELRPRTSPNTRSRSHPEPAIHPKRPQVDVDFSLPLAKVDELQOR
 LRQTDGSKHFPVSPRAOGKSVGVGVHVARGNKRPGLPAIPVSSPAAALP
 SLAPLGHKKHKAKESSQGCGRGLAPLASGCVLRLREHLRELPALVVVESQAGRPVQ
 RHEHNHNHNHNHNHNHNYGT"
 ORIGIN
 Query Match 77.4%; Score 1438; DB 9; Length 1438;
 Best Local Similarity 100.0%; Pred. No. 1.6e-250;
 Matches 1438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 217 CCCGACAGTGGGGAACCTTCACTCAAGCGCGCGCGTGTGCAAGCGAGGAGAGCGCC 276
 DB 1 CCCGACAGTGGGGAACCTTCACTCAAGCGCGCGCGTGTGCAAGCGAGGAGAGCGCC 60
 QY 277 GGAAGGTGACAGCTTGCCTGAGCGCTGCTGGGCTCGGAAGGAGATCGAGAGTGGAT 336
 DB 61 GGAAGGTGACAGCTTGCCTGAGCGCTGCTGGGCTCGGAAGGAGATCGAGAGTGGAT 120

Qy	337	CGGAGAGCAGCGCTGCCGGGCGGTGTCTCGGGAGCCCCGACAGCTGCGGTGGCGGGCAC	396
Db	121	CGGAGAGCAGCGCTGCCGGGCGGTGTCTCGGGAGCCCCGACAGCTGCGGTGGCGGGCAC	180
Qy	397	CATAGGCGGAAGCACCAGGAGCTCGTGGGCGACGTGTGGAGACAACGCTACGCGAGA	456
Db	181	CATAGGCGGAAGCACCAGGAGCTCGTGGGCGACGTGTGGAGACAACGCTACGCGAGA	240
Qy	457	AGAGGAGAGCAGCTTTGGCTGGAGGTGGCCCTGCTCTGAGAAACTGAACGGGCTGGG	516
Db	241	AGAGGAGAGCAGCTTTGGCTGGAGGTGGCCCTGCTCTGAGAAACTGAACGGGCTGGG	300
Qy	517	CAGCGAGATGAGAAAGATGAGAGATGAGCGAAACCTGCGCCAGAGCTCCAAAGACA	576
Db	301	CAGCGAGATGAGAAAGATGAGAGATGAGCGAAACCTGCGCCAGAGCTCCAAAGACA	360
Qy	577	GCTGAAGTTGAAGAGCTCCAGTGCAGCTGTCCATGAGAGAGACAGCCGGCAGAGTG	636
Db	361	GCTGAAGTTGAAGAGCTCCAGTGCAGCTGTCCATGAGAGAGACAGCCGGCAGAGTG	420
Qy	637	GACCTTCAACCTGTATGACTTTGACAAACAACGCGAAGTCAACCGAAGACATCACAG	696
Db	421	GACCTTCAACCTGTATGACTTTGACAAACAACGCGAAGTCAACCGAAGACATCACAG	480
Qy	697	CTTGCTGCAACCATCTATGAGGTGGTGGACTCCTCTGTCAACCACTCCCAACATCCAG	756
Db	481	CTTGCTGCAACCATCTATGAGGTGGTGGACTCCTCTGTCAACCACTCCCAACATCCAG	540
Qy	757	CAAGATCTCGGGTAAAGTCAACCGTGGCCCCGATGGCAGCCAGACAAAGAGAGCGT	816
Db	541	CAAGATCTCGGGTAAAGTCAACCGTGGCCCCGATGGCAGCCAGACAAAGAGAGCGT	600
Qy	817	CTTTGTCAATCAAGCTCACTGCAAGACCGCAAGGCCCGGAGAGACCAAGCCCACTGA	876
Db	601	CTTTGTCAATCAAGCTCACTGCAAGACCGCAAGGCCCGGAGAGACCAAGCCCACTGA	660
Qy	877	GGAAGCTGAGAGCTGGGAGAGAGAGCAGAGCCCGCTCAAGTTCAGGGTGAACAGCG	936
Db	661	GGAAGCTGAGAGCTGGGAGAGAGAGCAGAGCCCGCTCAAGTTCAGGGTGAACAGCG	720
Qy	937	CTTGAGCAGTCTGGCTGCTACACCATTTGGTATGAGAACTGCAAGAGAGAGAAACA	996
Db	721	CTTGAGCAGTCTGGCTGCTACACCATTTGGTATGAGAACTGCAAGAGAGAGAAACA	780
Qy	997	CTACTTATGATCTCGCCGGGATAGAAAATTAACGCTCCCAATTTGGGCTGCTCCCTTC	1056
Db	781	CTACTTATGATCTCGCCGGGATAGAAAATTAACGCTCCCAATTTGGGCTGCTCCCTTC	840
Qy	1057	CGTGGCCCAAGAGTCAAGACTGCCCGCCCGCACTCCCAATCCCACTGATCTGCTCCA	1116
Db	841	CGTGGCCCAAGAGTCAAGACTGCCCGCCCGCACTCCCAATCCCACTGATCTGCTCCA	900
Qy	1117	TGAGCCGGAAGCCATCCACATCCGACACGGAAGCCCGCAGGCGTGGACCCGGCCTCTT	1176
Db	901	TGAGCCGGAAGCCATCCACATCCGACACGGAAGCCCGCAGGCGTGGACCCGGCCTCTT	960
Qy	1177	CCAATTCTTGAACACCCCAATCGCAAGGTCTGAGAGTCCAGCAAGAGCTCCGGGAC	1236
Db	961	CCAATTCTTGAACACCCCAATCGCAAGGTCTGAGAGTCCAGCAAGAGCTCCGGGAC	1020
Qy	1237	CGAGAGCGGAGCAAGCACTTTGAGAGTCCCGCAAGGCCCGAGGCAAGAGTGGGTGT	1296
Db	1021	CGAGAGCGGAGCAAGCACTTTGAGAGTCCCGCAAGGCCCGAGGCAAGAGTGGGTGT	1080
Qy	1297	GGGCGACGTGGCCAGAGGGGCAAGAAAACAAGCCCTCTGAGGACCCGCACTCCCTGCGGT	1356
Db	1081	GGGCGACGTGGCCAGAGGGGCAAGAAAACAAGCCCTCTGAGGACCCGCACTCCCTGCGGT	1140
Qy	1357	GTCCCCCTCGGCCCAACCTGCTGCGAGCCCGGCTCTCTCCCTCTCTAGCCCTCCCTCGG	1416
Db	1141	GTCCCCCTCGGCCCAACCTGCTGCGAGCCCGGCTCTCTCCCTCTCTAGCCCTCCCTCGG	1200
Qy	1417	GCACAAAGACCAAGACCCAGCCAAAGAGAGAGCAGAGGGCTGCGGGGCTGCAAGGC	1476

Db	1201	GCACAAAGACCAACACCAAGACCCAAAGAGACCAAGAGGCTGCCGGGCTGCAAGGC	1260
Qy	1477	ACCACTGGCTCAGGTGGCCCTGTCTCGGGCGGAGACACTTCGCGGAGCTGCCGCTT	1536
Db	1261	ACCACTGGCTCAGGTGGCCCTGTCTCGGGCGGAGACACTTCGCGGAGCTGCCGCTT	1320
Qy	1537	GGTGGTATGAGAGCAGGCGGGGACCGGTCAGAGACATGAGACCAACCAACA	1596
Db	1321	GGTGGTATGAGAGCAGGCGGGGACCGGTCAGAGACATGAGACCAACCAACA	1380
Qy	1597	TGAACATCAACCAATTACCAACCACTTACCAAGACATAGAGCCCTCCCAAGGCC	1654
Db	1381	TGAACATCAACCAATTACCAACCACTTACCAAGACATAGAGCCCTCCCAAGGCC	1438

RESULT 5
 AR405917
 LOCUS AR405917 1438 bp DNA linear PAT 18-DEC-2003
 DEFINITION Sequence 5 from patent US 6630323.
 ACCESSION AR405917
 VERSION AR405917.1 GI:40154965
 KEYWORDS
 SOURCE
 ORGANISM
 Unknown.
 Unclassified.
 1 (bases 1 to 1438)
 REFERENCE
 Scott, M., Zeng, W. and Wharton, K.
 TITLE Naked cuticle genes and their uses
 JOURNAL Patent: US 6630323-A 5 07-OCT-2003;
 FEATURES
 source
 1..1438
 /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN
 Query Match 77.2%; Score 1434.8; DB 6; Length 1438;
 Best Local Similarity 99.9%; Pred. No. 6.1e-250;
 Matches 1436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	217	CCCCAGCATGGGAAACCTTCACTCAAGCCGGCCGCTGTGCAAGCGCAGGAGAGCCCC	276
Db	1	CCCCAGCATGGGAAACCTTCACTCAAGCCGGCCGCTGTGCAAGCGCAGGAGAGCCCC	60
Qy	277	GGAAGTGAACGCTTGCCTGTGAGCCCTGCTGGGCTCCGAAGGGCATGAGAGTGGAT	336
Db	61	GGAAGTGAACGCTTGCCTGTGAGCCCTGCTGGGCTCCGAAGGGCATGAGAGTGGAT	120
Qy	337	CGGAGACAGCGCTGCCCGGCGGTGTCTCGGACCCCGACAGCTGCGTTGGCGGGCAC	396
Db	121	CGGAGACAGCGCTGCCCGGCGGTGTCTCGGACCCCGACAGCTGCGTTGGCGGGCAC	180
Qy	397	CATAGCCGAAGCACCAGGAGCTCGTGGCGACGTGTTGAGAGACACGCTCAGCGAGA	456
Db	181	CATAGCCGAAGCACCAGGAGCTCGTGGCGACGTGTTGAGAGACACGCTCAGCGAGA	240
Qy	457	AGAGGAGAGAGACTTTGCGGTGGAAGTGGCCCTGCTCTGAGAACTGAACGGGCTGGG	516
Db	241	AGAGGAGAGAGACTTTGCGGTGGAAGTGGCCCTGCTCTGAGAACTGAACGGGCTGGG	300
Qy	517	CAGCGAGATGAGAAAGATGAGAGATGAGCGAAACCTGCGCCAGAGCTCCAAAGACA	576
Db	301	CAGCGAGATGAGAAAGATGAGAGATGAGCGAAACCTGCGCCAGAGCTCCAAAGACA	360
Qy	577	GCTGAAGTTGAAGAGCTCCAGTGCAGCTGTCCATGAGAGAGACAGCCGGCAGAGTG	636
Db	361	GCTGAAGTTGAAGAGCTCCAGTGCAGCTGTCCATGAGAGAGACAGCCGGCAGAGTG	420
Qy	637	GACCTTCAACCTGTATGACTTTGACAAACAACGCGAAGTCAACCGAAGACATCACAG	696
Db	421	GACCTTCAACCTGTATGACTTTGACAAACAACGCGAAGTCAACCGAAGACATCACAG	480
Qy	697	CTTGCTGCAACCATCTATGAGGTGGTGGACTCCTCTGTCAACCACTCCCAACATCCAG	756


```

Db      481  CTTCGTCACACCATATATAGAGTGTGGAGCTCTCTGTGTAACCACTCCCAACATCCAG 540
Qy      757  CAAGATGCTGGGGGTAAGCTCAACCGTGGCCCCGATGGCAAGCAAGAGAGCGT 816
Db      541  CAGAGTGTGGGGGTAAGCTCAACCGTGGCCCCGATGGCAAGCAAGAGAGCGT 600
Qy      817  CCTGTTCATCAGAGTGTGGAGCGGCAAGAGCGGCGGCAAGAGCGGCGTGA 876
Db      601  CCTGTTCATCAGAGTGTGGAGCGGCAAGAGCGGCGGCAAGAGCGGCGTGA 660
Qy      877  GGACCTGCGGAGGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 936
Db      661  GGAACCTGCGGAGGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Qy      937  CCTGAGCAGAGTGTGGTGTCTACCACTTGTGATAGAGAGAGAGAGAGAGAG 996
Db      721  CCTGAGCAGAGTGTGGTGTCTACCACTTGTGATAGAGAGAGAGAGAGAGAG 780
Qy      997  CTACTTATGATCTGTGGCGGAGTAAAGAACTACAGTCCCAATTTGGGCTTCCCTTC 1056
Db      781  CTACTTATGATCTGTGGCGGAGTAAAGAACTACAGTCCCAATTTGGGCTTCCCTTC 840
Qy      1057  CGTGGCCGAGAGTCAAGACTGGCCCCCGGACCTCCCAATCCCACTGATCTGCTCCCA 1116
Db      841  CGTGGCCGAGAGTCAAGACTGGCCCCCGGACCTCCCAATCCCACTGATCTGCTCCCA 900
Qy      1117  TGAGCCGGAAGCATCCATCCACACCGAAGGCCCAAGAGCGGTGAGCCGCTCTT 1176
Db      901  TGAGCCGGAAGCATCCATCCACACCGAAGGCCCAAGAGCGGTGAGCCGCTCTT 960
Qy      1177  CCACTTCTTGAACATCCCAATGGCCCAAGGTCTCAAGCTTCAGCAAGCGGTCCGGGCGAC 1236
Db      961  CCACTTCTTGAACATCCCAATGGCCCAAGGTCTCAAGCTTCAGCAAGCGGTCCGGGCGAC 1020
Qy      1237  CCAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1296
Db      1021  TCAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Qy      1297  GGGCCAGAGTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1356
Db      1081  GGGCCAGAGTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Qy      1357  GTTCCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1416
Db      1141  GTTCCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
Qy      1417  GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476
Db      1201  GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Qy      1477  ACCAAGTGGCTGAGTGGCTGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1536
Db      1261  ACCAAGTGGCTGAGTGGCTGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Qy      1537  GGTGTGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1596
Db      1321  GGTGTGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Qy      1597  TGAAGATCAACCAATTAACAACAATTCTAACAAGAGAGAGAGAGAGAGAGAG 1654
Db      1381  TGAAGATCAACCAATTAACAACAATTCTAACAAGAGAGAGAGAGAGAGAGAG 1438

```

```

RESULT 6
LOCUS      BC034838      1765 bp      mRNA      linear      ROD 29-JUN-2004
DEFINITION Mus musculus naked cuticle 1 homolog (Drosophila), mRNA (CDNA clone
ACCESSION BC034838
VERSION   BC034838.1   GI:22028144
KEYWORDS  MGC.
SOURCE      Mus musculus (house mouse)

```

```

ORGANISM      Mus musculus
REFERENCE      1 (bases 1 to 1765)
AUTHORS      Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F.,
Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Donald, M.F., Casavant, T.L.,
Schetz, T.E., Brownstein, M.J., Utsid, T.B., Toshynuk, S.,
Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,
Abrams, R.D., Mullaby, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shvachenko, Y.,
Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalys, D.E.,
Scherer, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL      12477932
PUBMED      2 (bases 1 to 1765)
REFERENCE      Strausberg, R.
AUTHORS      Direct Submission
TITLE      Submitted (29-JUN-2002) National Institutes of Health, Mammalian
JOURNAL      Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Marcelllo Bento Soares, Ph.D.
CDNA Library Preparation: M. Bento Soares, University of Iowa
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,
Andy Chan, Steve S. Chand, William Chow, Allison Cloutier, Ruth
Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,
Teika Olson, Diana Palmquist, Anca Petrescu, Anna Lisa Prahbu,
Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska,
Duane Smalys, Jeff Stott, Miranda Tsai, George Yang, Jackie
Schein, Asim Siddiqui, Rob Holt, Marco Marra.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Series: IRK Plate: 67 Row: 9 Column: 5
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 14456707.
FEATURES
source
1. 1765
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="MGC:40995 IMAGE:1429943"
/tissue_type="Mammary gland"
/clone_lib="Soares_mammary_gland_NBMGC"
/lab_host="DH10B"
/notes="Vector: pT73-Pac"
1. 1765
/gene="Nkd1"
/db_xref="LOCUSID:93960"
/db_xref="MGI:2135954"
138_1553
/gene="Nkd1"

```

```

/codon_start=1
/product="naked cuticle 1 homolog"
/protein_id="AAH34838.1"
/db_xref="GI:22028145"
/db_xref="LocustID:93960"
/db_xref="MG1:2135954"
/translation="MGRKSLPAVVCVKRREPSGDSFVNSAAMARKGIEMWIGRQRP
GSVSPQRLRLGTGVRGRRLVGDTSRLALGEDDDPLEVALPEKIDISJSGDE
RKRMELSEGGQSKLKELELCLVADVPDSQKFTLYLDPDNNGKRAITKPL
HTIYVNVSVNHSPTSLKLVKLVLCVDSQKFTLYLDPDNNGKRAITKPL
ELRGEKQKQRLAPRGQSDHEPCDCHNCVDENIRRNHYLDLAGIENTSOGPSS
PSVAKSELPPRLINPTDRSHSEPPAAHYPHRRPGCVDPGSHLLPPAKASLQOR
LRGTDDGSHNFRPSRPAQSKMGMGGAAGARSKPLVPTHTVSRHALATSPALL
RLRHHKHKHKKHRAHNSQSCGGLQPLAAGSTVNGREQVELPAVVTYSQAGQAV
TLEHFGHHKHHKHHRAHNSQ"

```

ORIGIN

Query Match	67.8%;	Score 1259.6;	DB 10;	Length 1765;
Best Local Similarity	84.3%;	Pred. No. 3.6e-218;		
Matches 1497;	Conservative 0;	Mismatches 254;	Indels 25;	Gaps 6;

[illegible][illegible]

RESULT 7			
AR405916			
LOCUS	AR405916	1731 bp	
DEFINITION	Sequence 3 from patent US 6630323.	DNA	linear
			PAT 18-DEC-2003

ACCESSION AR405916
 VERSION AR405916.1 GI:40154964
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (pages 1 to 1731)
 AUTHORS Scott, M., Zeng, W. and Wharton, K.
 TITLE Naked cuticle genes and their uses
 JOURNAL Patent: US 6630323-A 3 07-OCT-2003;
 FEATURES location/Qualifiers
 source 1..1731
 /organism="unknown"
 /mol_type="genomic DNA"
 ORIGIN
 Query Match 65.7%; Score 1221.4; DB 6; Length 1731;
 Best Local Similarity 83.9%; Pred. No 3.1e-211;
 Matches 1460; Conservative 0; Mismatches 256; Indels 25; Gaps 6;
 QY 82 CGGCGGAGAGCGCGTCCCGGCGCGCTCGGCTCGGCTCGGCGGCTGCTTGC 141
 Db 6 CCGCGGCGGAGCGCGTCCCGGCGCGCTTAAGCTCGGCTCGGCGGCGGACTG--TGA 63
 QY 142 GGAAG 201
 Db 64 GGAAG 121
 QY 202 TCCCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 261
 Db 122 CCGCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 177
 QY 262 GCGCAG 321
 Db 178 GCGCAG 237
 QY 322 CATCGAG 381
 Db 238 CATCGAG 297
 QY 382 GCGGTTGGCGGAG 441
 Db 298 GAGATTGGAG 357
 QY 442 CAGCTCAG 501
 Db 358 GCGCTCCGAG 417
 QY 502 GACTGAG 561
 Db 418 GATCGAG 477
 QY 562 AGGCTCAAG 621
 Db 478 GGGCTCAAG 537
 QY 622 CAGCCGAG 681
 Db 538 CAGCCGAG 597
 QY 682 AG 741
 Db 598 TGAAG 657
 QY 742 CTCGCCAAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 801
 Db 658 TTCGCCAAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 717
 QY 802 GAGCAAG 861
 Db 718 GAGTAAAG 777
 QY 862 GACCAAG 921

Db 778 CACCAAAACCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 837
 QY 922 CCAGAGTGAAG 981
 Db 838 CAGAGGAG 897
 QY 982 CGAG 1041
 Db 898 TGAAG 957
 QY 1042 GCGTGGCTCCCTTCGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1101
 Db 958 ACCGAGATCCCTTCGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1017
 QY 1102 TCGATCTCGCTCCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1161
 Db 1018 TCGCTCTCGCTCCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1077
 QY 1162 GAG 1221
 Db 1078 GAG 1137
 QY 1222 ACGGCTCCGAG 1281
 Db 1138 ACGGCTCCGAG 1197
 QY 1282 CAAGAGTGGTGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1341
 Db 1198 CAAGAGTGGTGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257
 QY 1342 CGCATCCCTGGAG 1401
 Db 1258 CACCAACATATGCTCCCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 1317
 QY 1402 CTTAGCCCCCTCGAG 1461
 Db 1318 CTTAGCCCCCTCGAG 1377
 QY 1462 CCGGAGCTGAG 1518
 Db 1378 CCGGAGCTGAG 1437
 QY 1519 GCGGAG 1578
 Db 1438 GAGGAG 1497
 QY 1579 TGAAG 1638
 Db 1498 CGAAG 1557
 QY 1639 CCGCTCCGAG 1698
 Db 1558 CC-----CAGAGAGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1603
 QY 1699 TATTATTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1758
 Db 1604 TATTATTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1663
 QY 1759 CACTTAATTATTTAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1818
 Db 1664 CACTTAATTATTTAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1723
 QY 1819 T 1819
 Db 1724 T 1724

RESULT 8
 AF358134
 LOCUS
 DEFINITION Mus musculus naked cuticle-1 (Nkd1) mRNA, complete cds.
 ACCESSION AF358134

QY 1544 TATGAGACGAGGCGGCGAGCCGCTCCAGAGCATGAGACGACGACCATGAACAT 1603
DB 1403 TAGAGAGCCAGGCTGGGCGGCGCTCCAGAGACGAGACCATCAACGAGACAT 1462
QY 1604 CACACCATTTACGACCACTTTCTACAGACATAGAGCCCTCCCGAGGCGCCACCTTGCC 1663
DB 1463 CACACCATTTATACGACCTTCTACAGCCCTAGACC-----CAGCAGGCTGCC 1511
QY 1664 ATATGAGAGACCCGACCCCGGACGACGACGACATTTATTTCTATTATTTATTTAT 1723
DB 1512 ACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1568
QY 1724 ATGATGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1783
DB 1569 ATGGCAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1628
QY 1784 TAGAGATCTATGAGAAACAGAGAACTAACTTTAT 1819
DB 1629 TAGAGACATGGAACACAGAACTAACTTTAT 1664

RESULT 9
AF343352 1416 bp mRNA linear ROD 29-MAR-2001
LOCUS Mus musculus Nkd mRNA, complete cds.
DEFINITION AF343352
ACCESSION AF343352 GI:13487304
VERSION AF343352.1 GI:13487304
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1416)
Yan, D., Wallingford, J.B., Sun, T.Q., Nelson, A.M., Sakanaka, C., Reinhard, C., Harland, R.M., Pantl, W.J. and Williams, L.T.
Cell autonomous regulation of multiple Dishevelled-dependent pathways by mammalian Nkd
Proc. Natl. Acad. Sci. U.S.A. 98 (7), 3802-3807 (2001)
2 (bases 1 to 1416)
Yan, D., Wallingford, J.B., Sun, T.-Q., Nelson, A.M., Sakanaka, C., Reinhard, C., Harland, R.M., Pantl, W.J. and Williams, L.T.
Direct Submission
Submitted (26-JAN-2001) Research, Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608, USA
LOCATION/Qualifiers
1. 1416
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
1. 1416
/note="Nkd; similar to Drosophila naked cuticle"
/codon_start=1
/product="Nkd"
/protein_id="AAK27485.1"
/db_xref="GI:13487305"
/translation="MGKLSKPAVCKKRRSPGDSFAVSAAMARKIEWIGRORCP
GVSGLROLAGTGRGTRELVLGEEDDEPLLEVALPEKIDISGSGE
KMERLSBPGQAKKOLKPEELQDVSEVDSHOEMFTLYDDNNKQVLRDITSL
HTLYEVLDSSVYNSPTSSKTLRYKLVYVARGSSKSVLEPHNIDLOSTRPATKAE
ELGWKKQAPLRFGQDSHLEQPDCTYHNVDSIKRNRYLDLAGIENTVSQFGES
PSVAQSELRPRPISNPKRSKSHPELAHNPARRPGVDPSFHLDTPEAKSELQOR
LRGTDGSKLFRVSPRAQKMGKMGHGAARSKRPVPTTHTVSPSAHLATPALP
TLAPLGHKKHKAKESSQASRGELGRPLAAGSGTVMGRGVRLPVVVVVSQAQAV
QRENHNHNHNHNHNHNFYQ"

ORIGIN
Query Match 58.0%; Score 1078.4; DB 10; Length 1416;
Best Local Similarity 85.6%; Pred. No. 2.6e-185;
Matches 1212; Conservative 0; Mismatches 201; Indels 3; Gaps 1;
QY 224 ATGGGAACTTCACTCCAGCGCGCGCGCTGTGCAAGCGGAGAGAGCCCGGAAGT 283

DB 1 ATGGGAACTTCACTCCAGCGCGCGCGCTGTGCAAGCGGAGAGAGCCCGGAAGT 60
QY 284 GACAGCTTCGCGTGAGCGCTGCCTGGGCTCGGAAGGCGATGAGAGTGAATGGAGA 343
DB 61 GACAGCTTCGCTGAAGCGCTGCTGGGCGAAGAAAGGCATCGAGAGTGAATGGAGA 120
QY 344 GACGCTTCGCGCGCTGCCTGGGCTCGGACAGCTGCCTGGGCTGGGCTGGGCTGGG 403
DB 121 CAGCGCTTCGCGCGCTGCCTGGGCTCGGACAGCTGCCTGGGCTGGGCTGGGCTGGG 180
QY 404 CGAAGCAGCGCGAGCTGCCTGGGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 463
DB 181 CGAGGAGCTCGGAG 240
QY 464 GACGCTTCGCGCTGCCTGGGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 523
DB 241 GACGAGCTTCGCGCTGCCTGGGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
QY 524 GATGAG 583
DB 301 GATGAG 360
QY 584 TTTGAG 643
DB 361 TTTGAG 420
QY 644 ACCCTGATGACTTTGACCAACAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 703
DB 421 ACTCTATGATGACTTTGACCAACAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 704 CACACCATTTATGAGTGTGAGCTCTCTGTCACCACTCTCCCAATCTCAGCAGATG 763
DB 481 CATACCATTTATGAGTGTGAGCTCTCTGTCACCACTCTCCCAATCTCAGCAGATG 540
QY 764 CTGCGGTTAAGCTCACCGGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 823
DB 541 CTGCGGTTAAGCTCACCGGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 824 AATCAGGCTGACCTGAG 883
DB 601 AATCAGGCTGACCTGAG 660
QY 884 CGAGCTGAG 943
DB 661 CGAGCTGAG 720
QY 944 CAGCTGCGCTGACCAACAG 1003
DB 721 CAGCTGCGCTGACCAACAG 780
QY 1004 GATCTGCGCGGATAGAAAATCAAGCTCCCAATTTGGGCTGCTCCCTTCGAGGCG 1063
DB 781 GACCTGCGCGGATAGAAAATCAAGCTCCCAATTTGGGCTGCTCCCTTCGAGGCG 840
QY 1064 CAGAGCTGAGAGCTGCGCGCGCGAGCTTCATCCATCTGATCTGCTCCATGAGCGG 1123
DB 841 CAGAGCTGAGAGCTGCGCGCGCGAGCTTCATCCATCTGATCTGCTCCATGAGCGG 900
QY 1124 GAGGCAATCCATCCACAG 1183
DB 901 GAGGCAATCCATCCACAG 960
QY 1184 CTTGACACCCCAATTCGCAAGGCTCTGAGAGCTCAGCAACGAGCTCGGAGGAGAC 1243
DB 961 CTTGACACCCCAATTCGCAAGGCTCTGAGAGCTCAGCAACGAGCTCGGAGGAGAT 1020
QY 1244 GGGAGCAAGCACTTTGTAGAGTCCCAAGGCTCAGGAGCAAGAGTGTGGTGGCCAC 1303
DB 1021 GGGAGCAAGCACTTTGTAGAGTCCCAAGGCTCAGGAGCAAGAGTGTGGTGGCCAC 1080
QY 1304 GTGCGCAGAGGAGGAG 1363

Db 1081 GGGGCCAGAGGTGCAAGAGCAAGCTTCACTGTATACCAACCACTATGTCCTCCCC 1140
QY 1364 TCGGCCACCTGGTGGCAAGCCGGCCCTCTCTCCCTCCCTAGCCCCCTCGGGCAAG 1423
Db 1141 TCTGCCATCTGGGCAACAGCCGAGCCCTTCTCCCACTCCCTGGAGCCCTGGGGCAAG 1200
QY 1424 AAGACAAGACCGAGGCAAGAGAGCAGAGAGGCTGGCGGGCTGCAAGGACCACTG 1483
Db 1201 AAACAACAAGATGAGAGCAAGAGAGCCAGAGAGCTGGCGGGCTTGCAGAGGCCCTTG 1260
QY 1484 GCCTCAGAGTGG---CCCTGTCTGGGGCGGAGAGCACTGCGAGAGCTGCCCGCCTTGATG 1540
Db 1261 GCTGCAAGAGGCTTCACTGTATGAGGGGAGAGGTGAGGGAGCTGCGCTGGCTGATG 1320
QY 1541 GTGTATAGAGCCAGAGCCGGGAGCCGCTTCAAGAGATATAGACCAACCAACATGAA 1600
Db 1321 GTGTACAGAGGCGAGGCTGGGAGAGCCGTCCAGAGACAGAAACCATCACCAAGAA 1380
QY 1601 CATCACCACTTATCACCACTTCTATCAGCCCTAG 1416
Db 1381 CATCACCACTTATCACCACTTCTATCAGCCCTAG 1416

RESULT 10
AX172351 1401 bp DNA linear PAT 03-JUL-2001
LOCUS AX172351
DEFINITION Sequence 1 from Patent WO0144279.
ACCESSION AX172351
VERSION AX172351.1 GI:14597503
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus (house mouse)
REFERENCE 1
AUTHORS Van, D. and Williams, L. T.
TITLE Mammalian dishevelled-associated proteins
JOURNAL Patent: WO 0144279-A 1 21-JUN-2001;
Chiron Corporation (US)
FEATURES
source 1. 1401
/organism="Mus musculus"
/mol_type="unassigned DNA"
/db_xref="taxon:10090"

ORIGIN
Query Match 57.4%; Score 1066.6; DB 6; Length 1401;
Best Local Similarity 85.6%; Pred. No. 3.6e-183;
Matches 1199; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY 224 ATGGGGAAATCTTCACTCCAGCCGGCGCGGTGTGCAAGCGAGAGAGAGCCGGAAAGT 283
Db 1 ATGGGGAAATCTTCACTCCAGCCGGCGCGGTGTGCAAGCGAGAGAGAGCCGGAAAGT 60
QY 284 GAGAGCTTGGCGGAGCGCTGCTGGGCTGGAAGGAGCATCGAGAGTGGATGGGAGA 343
Db 61 GAGAGCTTGGCGGAGCGCTGCTGGGCTGGAAGGAGCATCGAGAGTGGATGGGAGA 120
QY 344 CAGCGTGGCCGGGCGGTGTCTCGGAGCCCGGACAGCTGCGGTGGCGGGACCATAGAC 403
Db 121 CAGCGTGGCCGGGCGGTGTCTCGGAGCCCGGACAGCTGCGGTGGCGGGACCATAGAC 180
QY 404 CGAAGCAACCGGAGCTGCTGGGCGAGCGTGTGAGAGCACGCTCAGCGAGAGAGAG 463
Db 181 CGAAGCAACCGGAGCTGCTGGGCGAGCGTGTGAGAGCACGCTCAGCGAGAGAGAG 240
QY 464 GAGCACTTTTGGCTGGAAGTGGCCCTGCTCTGAGAGAGATGACAGGGCTGGGAGAG 523
Db 241 GAGCACTTTTGGCTGGAAGTGGCCCTGCTCTGAGAGAGATGACAGGGCTGGGAGAG 300
QY 524 GATGAGAAGAGATGAGAGAGTGAAGCAACCTGCCAGAGCTCAAGAAAGCATGTAAG 583
Db 301 GATGAGAAGAGATGAGAGAGTGAAGCAACCTGCCAGAGCTCAAGAAAGCATGTAAG 360

QY 584 TTGGAAGCTCCAGTGGCGAGCGTTCATGAGAGAGAGAGCCGGGAGAGTGAACCTTC 643
Db 361 TTGGAAGCTCCAGTGGCGAGCGTTCATGAGAGAGAGAGAGCCGGGAGAGTGAACCTTC 420
QY 644 ACCCTGTATGACTTTTGAACAACAAGGCAAGTCAACCGAGAGAGACATCACAGCTTGTG 703
Db 421 ACTCTATATGACTTTTGAACAACAAGTCAACCGGAGAGACATCACAGCTTGTG 480
QY 704 CACACCATCTATAGAGTGTGAGACTCTCTGTCAACCTTCCCAATCTCAGCAAGATG 763
Db 481 CATACCATCTATAGAGTGTGAGACTCTCTGTCAACCTTCCCAATCTCAGCAAGATG 540
QY 764 CTGGGGGTAAAGCTCAACCGTGGCCCGGATGGCAGGCAAGAGAGAGAGAGCTTGTG 823
Db 541 CTGGGGGTAAAGCTCAACCGTGGCTCTTGAACGAGGCAAGATGAGAGAGAGCTTGTG 600
QY 824 AATCAGGCTGACCTTGCAGAGCGCAAGGCCCGGAGCAAGACCAAGCCACTGAGAGACTG 883
Db 601 AATCAGGCTGACCTTGCAGAGCGCAAGGCCCGGAGCAAGACCAAGCCACTGAGAGACTG 660
QY 884 CGAGCTGGGAGAAAGAGAGAGCGAGCCCGCTCAGAGTTCAGAGGTGACAGCCCGCTGGAG 943
Db 661 CTGGGCTGGGAGAAAGAGAGCGAGCCCGCTCAGAGTTCAGAGGTGACAGCCCGCTGGAG 720
QY 944 CAGTCTGGCTGTACCAACCACTTGGATGAGAGAACTGAGAGAGAGAAACCACTACTTA 1003
Db 721 CAGCAGACTGTCTACCAACCACTTGGATGAGAGAACTTGAAGAGAGAAACCACTACTTA 780
QY 1004 GATCTGCGCGGATGAGAAATCTACAGTCCCAATTTGGGCTGAGCTCCCTTCCTGGGCC 1063
Db 781 GATCTGCGCGGATGAGAAATCTACAGTCCCAATTTGGGCTGAGCTCCCTTCCTGGGCC 840
QY 1064 CAGAGTCAAGACTGCCCCCGGCACTCTCAATCTCAGTGTGCTCTCCATGAGCCG 1123
Db 841 CAGAGTCAAGAGTGGCCCTTCAATCTCAACCCCACTGCTCTCTCCCAAGAGCA 900
QY 1124 GAAGCATCCACATCCCAACCGAAGGCCCAAGGCGTGAAGCCGGCTCTTCCACTTC 1183
Db 901 GAAGCATCCCAATCCCAACCGAGGCCCAAGGCTGAGAGCCCAAGCTCTTCTCACTTC 960
QY 1184 CTTGACACCCCAATCCCAATCCCAAGTCTCAGAGCTCCAGCAAGGCTCCGGGCAACCCAGAG 1243
Db 961 CTTGACACCCCAATTTTGGCAAGGATCAGAGACTCCAGCAAGGCTCCGGGCAACTCAAGAT 1020
QY 1244 GGGAGCAAGCACTTTGTAGGTCCCCAAGGCCCAAGGCAAGAGTGTGGTGGGCAC 1303
Db 1021 GGGAGCAAGCACTTTGTAGGTCCCCAAGGCCCAAGGCAAGAGTGTGGTGGGCAC 1080
QY 1304 GTGGCCAGAGGCGAAGAAACAAGCCCTCTGGGAGCCGSCAATCCCTGGGAGTCCCC 1363
Db 1081 GGGCCAGAGGTGCAAGAAACAAGCTTCACTGATCCCAACCCCAATCTGTCTCCCC 1140
QY 1364 TCGGCCACCTGGCTGCAAGCCGGCCCTCTCTCCCTCCCTAGCCCCCTCGGGCAAG 1423
Db 1141 TCTGCCATCTGGGCAACAGCCAGCCCTTCTCCCACTTGGCAACCCCTGGGGCAAG 1200
QY 1424 AAGACAAGACCGAGGCAAGAGAGCAGAGAGGCTGGCGGGCTGCAAGGACCACTG 1483
Db 1201 AAACAACAAGATGAGAGCAAGAGAGCCAGAGAGCTGGCGGGCTTGCAGAGGCCCTTG 1260
QY 1484 GCCTCAGAGTGG---CCCTGTCTGGGGCGGAGAGCACTGCGGAGAGCTGCCCGCCTTGATG 1540
Db 1261 GCTGCAAGAGGCTTCACTGTATGAGGGGAGAGAGGTGAGGGAGCTGCTGGCTGATG 1320
QY 1541 GTGTATAGAGCCAGAGCCGGGAGCCGCTTCAAGAGATGAGAGCAACCAACATGAA 1600
Db 1321 GTGTACAGAGGCGAGGCTGGGAGAGCCGTCCAGAGACAGAAACCATCACCAAGAA 1380
QY 1601 CATCACCACTTATCACCACTTCTATCAGCCCTAG 1621
Db 1381 CATCACCACTTATCACCACTTCTATCAGCCCTAG 1401

[illegible]

QY	1136	ATCCACACCGAAAGCCCCCAAGGGGTGAGACCCGGGCTCTTCACATTCCTTGAACCCCA	1135
Db	736	ATCCACACCGAAAGCCCCCAAGGGGTGAGACCCGGGCTCTTCACATTCCTTGAACCCCA	795
QY	1196	ATCCGCAAGGTCTCAGAGCTTCAGCAACGGCTCCGGGGACCCAGAGACGGAGGCAAGCAC	1255
Db	796	ATCCGCAAGGTCTCAGAGCTTCAGCAACGGCTCCGGGGACCCAGAGACGGAGGCAAGCAC	855
QY	1256	TTTGTGAGGTCTCCCAAGGCCCAAGGCAAGAGTGTGGGTGTGGGCCACGTGACAGAGG	1315
Db	856	TTTGTGAGGTCTCCCAAGGCCCAAGGCAAGAGTGTG-----	891
QY	1316	GCAAGAAACAGGCCCTCTGGAGACCCGGCCATCTCTGGGTGTCCCTCCGCCACCTG	1375
Db	892	-----	891
QY	1376	GCTGCAGACCCGGGCTCTCTCCCTCCCTAGCCCCCTCGGGACAGAAAGCACAGACAC	1435
Db	892	-----GCCCTCTCGGGACAGAAAGCACAGACAC	921
QY	1436	CGAGCCAAAGAGAGACCAAGCAGGAGTCCCGGGGCTTCAGAGGACCATCTGCTTCAAGTGGC	1495
Db	922	CGAGCCAAAGAGAGACCAAGCAGGAGTCCCGGGGCTTCAGAGGACCATCTGCTTCAAGTGGC	981
QY	1496	CCTGTCCTGGGGGGGAGGACCACTGTCGGGAGCTGCGCGCTTGAGTGTATGAGAGCGAG	1555
Db	982	CCTGTCCTGGGGGGGAGGACCACTGCGGAGCTGCGCGCTTGAGTGTATGAGAGCGAG	1041
QY	1556	GCCGGGAGCGCG	1567
Db	1042	GCCGGGAGCGCG	1053
RESULT 12			
LOCUS	HSA303140	163119 bp	DNA
DEFINITION	Homo sapiens NOD2 gene for LRR-containing protein, exon 1-11.	linear	PRI 03-APR-2004
ACCESSION	AIJ03140		
VERSION	AJ303140.1	GI:14488148	
KEYWORDS	LRR-containing protein; NOD2 gene.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 Hugot,J.P., Chamaillard,M., Zouali,H., Lesage,S., Cezard,J.P., Belaiche,J., Almer,S., Tysk,C., O'Morain,C.A., Gassull,M., Binder,V., Finkel,Y., Cortot,A., Modigliani,R., Laurent-Puig,P., Gower-Rousseau,C., Macry,J., Colombel,J.F., Sahbatou,M. and Thomas,G.		
TITLE	Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease		
JOURNAL	Nature 411 (6837), 599-603 (2001)		
MEDLINE	21279172		
PUBMED	11385576		
REFERENCE	2 (bases 1 to 163319)		
AUTHORS	Zouali,H.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-JAN-2001) Zouali H., Genetics, Fondation Jean Dausset-CEPH, 27 RUE Juliette Dodu, 75010 Paris France, FRANCE		
FEATURES	Location/Qualifiers		
source	1..163319		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="16"		
	/map="16q12-13"		
	/clone="hb87b10"		
	/clone_1lb="CEPH Human BAC library"		
	98302..131954		
	/gene="NOD2"		
	join(98302..98906,106807..106912,109510..111325,		

115517. .115600,115821. .115904,118855. .118938.
121552. .121635,122230. .122313,124418. .124501,
128745. .128828,130674. .131994)
/gene="NOD2"
98302. .98906
/gene="NOD2"
/number=1
join(98448. .98906,106807. .106912,109510. .11325,
115517. .115600,115821. .115904,118855. .118938,
121552. .121635,122230. .122313,124418. .124501,
128745. .128828,130674. .130746)
/gene="NOD2"
/function="putative NF-kB activator"
/note="protein of the CARD family"
/codon_start=1
/product="LRR-containing protein"
/protein_id="CAC42117.1"
/db_xref="GI:14488149"
/db_xref="GOA:Q9HC29"
/db_xref="Swiss-Prot:Q9HC29"
/translacion="MCSQEAFOAQRSQLVELLVSGSLGSPESVLDMVLMSWEVLISWEDY
EFPHLIGDPLSHLRRLLDYTNKGTACOKLIAAOEAQDOSPKHGCDMDSHLH
PRDIOGRPAIVRLSHVENMLDIAERGVISOECDELRLPIFPGSARRLDLD
ATVKNGLAARPLIQHVQSLPYPLALPLEAATCKYMAKLRPTVSAQSRFLSTVGAATP
LLELDTYBENVLEVWADVGMAGSPQKSPATIGLELSTPGHLDDADDTIVIVSAGSP
GSTFLPRLHLLMAAGDOFOEFLFVPPSCRLQCMAPLSVRTLLEPHCCWPDVGQD
DIFOLLHPRDFVLTLTDFGDFEFKFRPFDREHSGPTDPTSVQTLFNLQGNLKNNA
RVYVSRPAKFIKFIHPIPAAPGEAKSVIAMPQFTWLRSLYEMOEBERLAKVAGLVNHH
KLTSPVSGTEGCAALAVLQHLRPAVALQIDYNSVDIGVQLPCLVCAKALVRDNGH
LMSDGIKLCIECALHGOLOKTLFNNKLDGCHSAWAKLACRONMLAIRDNNTYIT
TAAGQVLAIEKRGNTSIQFLGFMKRGVDEGAOLMALADHOSLRHLSVGNNTSS
VQAQALMLANVLEIECLSENLQDEGCSLAEGKKNSSLKTLKLSNNCTITGLG
AEALQALRENDTILVWLRGNTFLSEVDKLCGRDRLI,"
98907. .106806
/gene="NOD2"
/number=1
106807. .106912
/gene="NOD2"
/number=2
106913. .109509
/gene="NOD2"
/number=2
109510. .11325
/gene="NOD2"
/number=3
11326. .11516
/gene="NOD2"
/number=3
11517. .115600
/gene="NOD2"
/number=4
115601. .115820
/gene="NOD2"
/number=4
115821. .115904
/gene="NOD2"
/number=5
115905. .11854
/gene="NOD2"
/number=5
11855. .118938
/gene="NOD2"
/number=6
118939. .121551
/gene="NOD2"
/number=6
121552. .121635
/gene="NOD2"

intron	/number=7 121636..122229 /gene="NOD2"
exon	/number=7 122230..122313 /gene="NOD2"
intron	/number=8 122314..124417 /gene="NOD2"
exon	/number=8 124418..124501 /gene="NOD2"
intron	/number=9 124502..128744 /gene="NOD2"
exon	/number=9 128745..128828 /gene="NOD2"
intron	/number=10 128829..130673 /gene="NOD2"
exon	/number=10 130674..131994 /gene="NOD2"
	/number=11

Query Match	42.1%	Score 783	DB 9	Length 163319
Best Local Similarity	100.0%	Pred. No. 1,2e-131	Mismatches 0	Indels 0
Matches 783	Conservative 0	Midmatches 0	Gaps 0	
Qy 1046	GGCTCCCTTCGCGGCGCCAGAAAGTCAGAACTGCGCCCGCGCACCTTCAAATCCCACTCGA	1105		
Db 31866	GGCTCCCTTCGCGGCGCCAGAAAGTCAGAACTGCGCCCGCGCACCTTCAAATCCCACTCGA	3202		
Qy 1106	TCTGCTCCCATGAGCGGAAAGCCATCCATCCCAACCGAAAGCCCAAGGCGTGAGAC	1165		
Db 32026	TCTGCTCCCATGAGCGGAAAGCCATCCATCCCAACCGAAAGCCCAAGGCGTGAGAC	3208		
Qy 1166	CCGGCCCTTCCTCCCACTTCCTTGACACCCCAATGCGCAAGGCTTCAGAGCTCCAGCAACGG	1225		
Db 32086	CCGGCCCTTCCTCCCACTTCCTTGACACCCCAATGCGCAAGGCTTCAGAGCTCCAGCAACGG	3214		
Qy 1226	CTCGGGGCGAACCCAGAGCGGAGCAACACTTTGTAGAGTCTCCCGCAAGGCCAGAGGCAAG	1285		
Db 32146	CTCGGGGCGAACCCAGAGCGGAGCAACACTTTGTAGAGTCTCCCGCAAGGCCAGAGGCAAG	3220		
Qy 1286	AGTGTGGGTGTGGGCCACGTGGCCAGAGGGGCAAGAAACAAGCCCTCTTGGACCGGCC	1345		
Db 32206	AGTGTGGGTGTGGGCCACGTGGCCAGAGGGGCAAGAAACAAGCCCTCTTGGACCGGCC	3226		
Qy 1346	ATCCCTCGCGGTGTCCTCCCTCCGCCCACTGGCTGGCGCAGCCCGCGCCCTCCCTCCCTCA	1405		
Db 32266	ATCCCTCGCGGTGTCCTCCCTCCGCCCACTGGCTGGCGCAGCCCGCGCCCTCCCTCCCTCA	3232		
Qy 1406	GCCTCCCTCGGGGCAAGAAAGCAACAGCACCGAGCCAGAGGAGACCCAGAGGGCTGCGGG	1465		
Db 32326	GCCTCCCTCGGGGCAAGAAAGCAACAGCACCGAGCCAGAGGAGACCCAGAGGGCTGCGGG	3238		
Qy 1466	GGCTCGAGGCAACACTGGCTCAAGTGGCCCTGTCTTGGGGGGGGAGCACTTGGCGGAG	1525		
Db 32386	GGCTCGAGGCAACACTGGCTCAAGTGGCCCTGTCTTGGGGGGGGAGCACTTGGCGGAG	3244		
Qy 1526	CTGCGCCGCTTTGGTGTGTATGAGACGCGAGCGGGGCGAGCCGGTCCAGAGCATGAGCAC	1585		
Db 32446	CTGCGCCGCTTTGGTGTGTATGAGACGCGAGCGGGGCGAGCCGGTCCAGAGCATGAGCAC	3250		
Qy 1586	CACCAACCACTGAAATCACCACTATTACCACTTTACACGAGCATAGAGCCCTCC	1645		
Db 32506	CACCAACCACTGAAATCACCACTATTACCACTTTACGAGCATAGAGCCCTCC	3256		
Qy 1646	CGAGGGCCCACTGCGCATGTGAAGACCCCACTCCGAGACCAACAAGCATTTATT	1705		

Db 32566 CCAAGGCCCAACCCCTGCAATATGAAGAGACCCCAACCCCGACACCAAGCATTATTAAT 32625
QY 1706 CTATTAATTATTTGTTATTAATGATGATTAATTTGTTATTAATTAATTTGTTACTCCACTAAT 1765
Db 32626 CTATTAATTATTTGTTATTAATGATGATTAATTTGTTATTAATTAATTTGTTACTCCACTAAT 32685
QY 1766 ATTTAGCTAGCTACATGTAAGATCTATGGAACACAGAACTTAATTTATTTAT 1825
Db 32686 ATTTAGCTAGCTACATGTAAGATCTATGGAACACAGAACTTAATTTATTTAT 32745
QY 1826 GTT 1828
Db 32746 GTT 32748
RESULT 13
LOCUS AC007608/c 169773 bp DNA linear PRI 19-MAR-2003
DEFINITION Homo sapiens chromosome 16 clone RP11-401P9, complete sequence.
ACCESSION AC007608
VERSION AC007608.7 GI:29124038
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 169773)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 169773)
Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
Jones,M., Bucknham,J., Chasteen,L., Thompson,S., Goodwin,L.,
Bryant,D., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
Campbell,C., Fawcett,J., Malbie,M., Bussod,M., Sutherland,R.,
McMurry,K., Han,C. and Deaven,L.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-1999) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
REFERENCE 3 (bases 1 to 169773)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (04-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 169773)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (11-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 5 (bases 1 to 169773)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Submitted (19-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Mar 19, 2003 this sequence version replaced gi:18129388.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www-bhg.scf.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
Location/Qualifiers
1.169773
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-401P9"
ORIGIN

Query Match 42.1%; Score 783; DB 9; Length 169773;
Best Local Similarity 100.0%; Pred. No. 1,2e-131;
Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1046 GGCCTCCCTCCCGGAGCCGAGAGTCAAGACTGCCCCCGGACCTCCATCCACTCGA 1105
Db 78524 GGCCTCCCTCCCGGAGCCGAGAGTCAAGACTGCCCCCGGACCTCCATCCACTCGA 78465
QY 1106 TCTGCTCCCATGAGCCGGAAGCATCCATCCACACCGAAAGCCCAAGGCTGAGC 1165
Db 78464 TCTGCTCCCATGAGCCGGAAGCATCCATCCACACCGAAAGCCCAAGGCTGAGC 78405
QY 1166 CCGGCTCTCTTCACTCTTCTTGAACCCCAATGCCCAAGCTTCAGCTTCAGAACGG 1225
Db 78404 CCGGCTCTCTTCACTCTTCTTGAACCCCAATGCCCAAGGCTTCAGAGCTTCAGAACGG 78345
QY 1226 CTCGGGGGACCCAGAGCGGGAAGCACTTTGTGTGTTGTCCTCCCAAGGCGGAGGAG 1285
Db 78344 CTCGGGGGACCCAGAGCGGGAAGCACTTTGTGTGTTGTCCTCCCAAGGCGGAGGAG 78285
QY 1286 AGTGTGTGTGTGAGCCACGTGTGCGCAGAGGGGCAAGAAACAGCCCTCTGGAACCCGCC 1345
Db 78284 AGTGTGTGTGTGAGCCACGTGTGCGCAGAGGGGCAAGAAACAGCCCTCTGGAACCCGCC 78225
QY 1346 ATCCCTGCGGTGTGTCCTCTCTGCGCCACCTGCTGCGCAGCCCGGCTCTCTCTCTTA 1405
Db 78224 ATCCCTGCGGTGTGTCCTCTCTGCGCCACCTGCTGCGCAGCCCGGCTCTCTCTCTTA 78165
QY 1406 GCGCCCTCGGGGCAAGAAAGCAAGCAACCGGCAAGGAGGAGGCGGAGGCTGCGG 1465
Db 78164 GCGCCCTCGGGGCAAGAAAGCAAGCAACCGGCAAGGAGGAGGCGGAGGCTGCGG 78105
QY 1466 GGCCTGAGGACACACCTGCTGAGTGTGAGCGGCGGAGGCGGAGGAGGAGGAGGAGG 1525
Db 78104 GGCCTGAGGACACACCTGCTGAGTGTGAGCGGCGGAGGCGGAGGAGGAGGAGGAGG 78045
QY 1526 CTGCTGCGGCTGT 1585
Db 78044 CTGCTGCGGCTGT 77985
QY 1586 CACCACACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1645
Db 77984 CACCACACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 77925
QY 1646 CCAAGGCCCAACCCCTGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1705
Db 77924 CCAAGGCCCAACCCCTGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 77865
QY 1706 CTATTAATTATTTGTTATTAATGATGATTAATTTGTTATTAATTAATTTGTTACTCCACTAAT 1765
Db 77864 CTATTAATTATTTGTTATTAATGATGATTAATTTGTTATTAATTAATTTGTTACTCCACTAAT 77805
QY 1766 ATTTAGCTAGCTACATGTAAGATCTATGGAACACAGAACTTAATTTATTTAT 1825
Db 77804 ATTTAGCTAGCTACATGTAAGATCTATGGAACACAGAACTTAATTTATTTAT 77745
QY 1826 GTT 1828
Db 77744 GTT 77742
RESULT 14
LOCUS AC007334/c 170791 bp DNA linear PRI 13-MAR-2003
DEFINITION Homo sapiens chromosome 16 clone RP11-147B17, complete sequence.
ACCESSION AC007334
VERSION AC007334.8 GI:28933524
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 170791)
 AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los
 TITLE Alamos National Laboratory.
 JOURNAL Direct Submission
 AUTHORS Unpublished
 2 (bases 1 to 170791)
 Bruce, D., Mundy, M., Doggett, N., Munk, C., Saunders, E., Robinson, D.,
 Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,
 Bryant, J., Tesmer, J., Meincke, L., Longmire, J., White, S., Tetun, O.,
 Campbell, C., Fawcett, V., Malbise, M., Buesod, M., Sutherland, R.,
 McMurtry, K., Han, C. and Deaven, L.
 TITLE Direct Submission
 JOURNAL Submitted (21-APR-1999) Center for Human Genome Studies, DOE Joint
 Genome Institute, Los Alamos National Laboratory, MS M888, Los
 Alamos, NM 87545, USA
 3 (bases 1 to 170791)
 TITLE Direct Submission
 JOURNAL Submitted (24-AUG-2001) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 4 (bases 1 to 170791)
 DOE Joint Genome Institute, Stanford Human Genome Center and Los
 Alamos National Laboratory.
 TITLE Direct Submission
 JOURNAL Submitted (13-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 On Mar 13, 2003 this sequence version replaced gi:15284265.
 COMMENT Draft Sequence Produced By DOE Joint Genome Institute
 www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center and Los Alamos
 National Laboratory
 www.snhgc.stanford.edu
 Quality: Phrap Quality >=40 100% of Sequence;
 Estimated total Number of Errors is 0.
 Location/Qualifiers
 1. 170791
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-147B17"

ORIGIN
 Query Match 42.1% Score 783; DB 9; Length 170791;
 Best Local Similarity 100.0%; Pred. No. 1.2e-131;
 Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1046 GGCCTCCCTTCCTGCGCCAGAACTGAGAACTGCCCCCGCACTCCCAATCCCACTCGA 1105
 DB 28069 GGCCTCCCTTCCTGCGCCAGAACTGAGAACTGCCCCCGCACTCCCAATCCCACTCGA 28010
 QY 1106 TCTGCTCCCATGAGCGGGAAGCATTCACATCCCAACGGAAGGCCCAAGCGTGGAC 1165
 DB 28009 TCTGCTCCCATGAGCGGGAAGCATTCACATCCCAACGGAAGGCCCAAGCGTGGAC 27950
 QY 1166 CCGGCTCCTTCCTGCACTTGAACGCCCATGCGCAAGGCTCAGAGCTCCAGCAACG 1225
 DB 27949 CCGGCTCCTTCCTGCACTTGAACGCCCATGCGCAAGGCTCAGAGCTCCAGCAACG 27890
 QY 1226 CTCGGGGGCAACCAAGAGCGGAGCACTTTGTGAGTTCCTCCCAAGCGCCAGGAG 1285
 DB 27889 CTCGGGGGCAACCAAGAGCGGAGCACTTTGTGAGTTCCTCCCAAGCGCCAGGAG 27830
 QY 1286 AGTGTGGGTGTGGGCAAGTGTGGGCAAGAGGGAAGAAACAAGCCCTCTGTGGAGCCG 1345
 DB 27829 AGTGTGGGTGTGGGCAAGTGTGGGCAAGAGGGAAGAAACAAGCCCTCTGTGGAGCCG 27770
 QY 1346 ATCCCTGCGGTGTGGGCAAGTGTGGGCAAGAGGGAAGAAACAAGCCCTCTGTGGAGCCG 1405
 DB 27769 ATCCCTGCGGTGTGGGCAAGTGTGGGCAAGAGGGAAGAAACAAGCCCTCTGTGGAGCCG 27710
 QY 1406 GCCCGCTCGGGGCAAGAGGCAAGAGCAACGAGCCCAAGAGAGACCGAGAGGCGTGGCG 1465

DB 27709 GCCCGCTCGGGGCAAGAGGCAAGAGCAACGAGCCCAAGAGAGACCGAGGCGTGGCG 27650
 QY 1466 GGCCTGAGGCAACCACTGGCTGAGGCTGCTGCTGGGGGGGAGCACTGGGGAG 1525
 DB 27649 GGCCTGAGGCAACCACTGGCTGAGGCTGCTGCTGGGGGGGAGCACTGGGGAG 27590
 QY 1526 CTCGGCGCTTGT 1585
 DB 27589 CTCGGCGCTTGT 27530
 QY 1586 CACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1645
 DB 27529 CACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 27470
 QY 1646 CCAAGGCGCCCAACCTGCAATATGAAGAGCCCAACCCCGACCAACCAAGGCAATATAT 1705
 DB 27469 CCAAGGCGCCCAACCTGCAATATGAAGAGCCCAACCCCGACCAACCAAGGCAATATAT 27410
 QY 1706 CTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1765
 DB 27409 CTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 27350
 QY 1766 ATTAGCTACCTTACATGTGAAAGATCTATGAAACACAGAACTAACTTTATTTATAT 1825
 DB 27349 ATTAGCTACCTTACATGTGAAAGATCTATGAAACACAGAACTAACTTTATTTATAT 27290
 QY 1826 GTT 1828
 DB 27289 GTT 27287

RESULT 15

AC145238 178790 bp DNA linear HTG 15-AUG-2003
 LOCUS AC145238
 DEFINITION Pan troglodytes clone CH251-568P19, WORKING DRAFT SEQUENCE, 25
 ordered pieces.

AC145238
 AC145238.2 GI:33667125
 VERSION HTG; HTGS PHASE2; HTGS_DRAFT.
 KEYWORDS Pan troglodytes (chimpanzee)
 SOURCE Pan troglodytes
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

REFERENCE

AUTHORS

1 (bases 1 to 178790)
 Antonellis, A., Ayala, K., Beckstrom-Stenberg, S.M., Benjamin, B.,
 Blakeley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Cariaga, K.,
 Chu, G., Coleman, B., Coleman, H., Engle, J., Granite, S., Guan, X.,
 Gupta, J., Haghighi, P., Han, J., Hansen, N., Ho, S.-D., Hu, P.,
 Hurle, B., Idol, J.R., Karlins, E., Kwong, P., Latic, P., Lee-Lin, S.-Q.,
 Legaspi, R., Maduro, Q.L., Maduro, V.B., Marquis, B.H., Masello, C.,
 Maskeri, B., McDowell, J., Paguirigan, C., Pearson, R., Portnov, M.E.,
 Prasad, A., Redix-Dugue, N., Schendler, K., Scheller, M.G., Shah, K.,
 Sloan, C., Stantrid, S., Thomas, J.W., Thomas, P.V., Tsipouri, V.,
 Vogt, J.L., Wetherby, K.D., Young, A. and Green, E.D.
 TITLE NISC Comparative Sequencing Initiative
 JOURNAL Unpublished
 2 (bases 1 to 178790)
 Green, E.D.

REFERENCE

AUTHORS

1 (bases 1 to 178790)
 TITLE Direct Submission
 JOURNAL Submitted (18-JUN-2003) NIH Intramural Sequencing Center, 8717
 Grovemont Circle, Gaithersburg, MD 20877, USA
 3 (bases 1 to 178790)
 Green, E.D.

REFERENCE

AUTHORS

1 (bases 1 to 178790)
 TITLE Direct Submission
 JOURNAL Submitted (15-AUG-2003) NIH Intramural Sequencing Center, 8717
 Grovemont Circle, Gaithersburg, MD 20877, USA
 On Aug 15, 2003 this sequence version replaced gi:31880081.

COMMENT

----- Genome Center
 Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: http://www.nisc.nih.gov
 Contact: nisc_zoo@ngi.nih.gov
 Project Information

Center project name: esk
Center clone name: 568P19

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics -----

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171494 bases at least Q40
Consensus quality: 174120 bases at least Q30
Consensus quality: 175495 bases at least Q20
Insert size: 149000; agarose-fp
Insert size: 176390; sum-of-contigs
Quality coverage: 10.92x in Q20 bases; agarose-fp
Quality coverage: 9.22x in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently consists of 25 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

This sequence will be replaced
by the finished sequence as soon as it is available and
the accession number will be preserved.

```
1 2604: contig of 2604 bp in length
* 2605 2704: gap of unknown length
* 2705 2790: contig of 5086 bp in length
* 7791 7890: gap of unknown length
* 7891 12833: contig of 4943 bp in length
* 12834 12933: gap of unknown length
* 12934 14664: contig of 1731 bp in length
* 14665 14764: gap of unknown length
* 14765 32237: contig of 17473 bp in length
* 32238 32337: gap of unknown length
* 32338 35356: contig of 3019 bp in length
* 35357 35456: gap of unknown length
* 35457 39292: contig of 3836 bp in length
* 39293 39392: gap of unknown length
* 39393 41210: contig of 1818 bp in length
* 41211 41310: gap of unknown length
* 41311 43762: contig of 2452 bp in length
* 43763 43863: gap of unknown length
* 43863 46766: contig of 2904 bp in length
* 46767 46866: gap of unknown length
* 46867 53405: contig of 6539 bp in length
* 53406 53505: gap of unknown length
* 53506 55334: contig of 1829 bp in length
* 55335 55434: gap of unknown length
* 55435 69095: contig of 13661 bp in length
* 69096 69195: gap of unknown length
* 69196 78811: contig of 9616 bp in length
* 78812 78911: gap of unknown length
* 78912 89715: contig of 10804 bp in length
* 89716 89815: gap of unknown length
* 89816 97259: contig of 7444 bp in length
* 97260 97359: gap of unknown length
* 97360 99238: contig of 1879 bp in length
* 99239 99338: gap of unknown length
* 99339 101324: contig of 1986 bp in length
* 101325 101424: gap of unknown length
* 101425 112068: contig of 10644 bp in length
```

FEATURES

SOURCE

```
* 112069 112168: gap of unknown length
* 112169 118679: contig of 6511 bp in length
* 118680 118779: gap of unknown length
* 118780 120680: contig of 1901 bp in length
* 120681 120780: gap of unknown length
* 120781 135733: contig of 14952 bp in length
* 135733 135833: gap of unknown length
* 135833 138724: contig of 2891 bp in length
* 138724 138823: gap of unknown length
* 138823 163652: contig of 24829 bp in length
* 163653 163753: gap of unknown length
* 163753 178790: contig of 15038 bp in length.
```

Location/Qualifiers

```
1. 178790
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="CH251-568P19"
/clone_1b="CH251"
1. 90615
/clone="clone overlaps with GenBank Accession Number
AC145175 clone CH251-32B1 (center project name esj)"
1. 2604
/clone="assembly_fragment"
vector_end="SP6"
vector_start="left"
2705..7790
/clone="assembly_fragment"
7891..12833
/clone="assembly_fragment"
12934..14664
/clone="assembly_fragment"
14765..32237
/clone="assembly_fragment"
32338..35356
/clone="assembly_fragment"
35457..39292
/clone="assembly_fragment"
39393..41210
/clone="assembly_fragment"
41311..43762
/clone="assembly_fragment"
43863..46766
/clone="assembly_fragment"
46867..53405
/clone="assembly_fragment"
53506..55334
/clone="assembly_fragment"
55435..69095
/clone="assembly_fragment"
69196..78811
/clone="assembly_fragment"
78912..89715
/clone="assembly_fragment"
89816..97259
/clone="assembly_fragment"
97360..99238
/clone="assembly_fragment"
99339..101324
/clone="assembly_fragment"
101425..112068
/clone="assembly_fragment"
112169..118679
/clone="assembly_fragment"
118780..120680
/clone="assembly_fragment"
120781..135732
/clone="assembly_fragment"
135833..138723
/clone="assembly_fragment"
138824..178790
/clone="clone overlaps with GenBank Accession Number
AC145174 clone CH251-260A1 (center project name esj)"
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM, nucleic - nucleic search, using sw model

Run on: December 29, 2004, 13:06:18 ; Search time 1305 Seconds
(without alignments)
7477.911 Million cell updates/sec

Title: US-09-993-966-5

Perfect score: 1859

Sequence: 1 gaattcgccctctatacgcg.....aaaaaaaaaagcgccgcg 1859

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 413486 seqs, 2624710521 residues

8269772

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1859	100.0	1859	10	AAL51511 Human Nkd
2	1434.8	77.2	1438	3	AAAG63925 DNA encod
3	1221.4	65.7	1731	3	AAAG63924 DNA encod
4	1078.4	58.0	1416	4	AAH28343 Nucleotid
5	1066.6	57.4	1401	10	AAL51512 Mouse Nkd
6	783	42.1	2379	4	AAL03214 Human rep
7	783	42.1	2379	4	AAL03216 Human rep
8	590	31.7	590	10	AAL51525 Human Nkd
9	455.8	24.5	1561	5	AAAG2082 DNA encod
10	418.2	22.5	427	5	ACH21474 Human adu
11	284.2	15.3	1743	5	AAAG78752 DNA encod
12	284.2	15.3	1743	5	AAAG2081 DNA encod
13	284.2	15.3	1743	10	ADCG1966 Human nov
14	272.4	14.7	1966	4	AAAG27117 cDNA enco
15	272.4	14.7	1966	10	ADBG3295 Human cDN
16	252	13.6	1285	3	AAAG63926 DNA encod
17	245.4	13.2	1885	10	ADFG2455 Leukaemia
18	215	11.6	215	10	AAL51526 Human Nkd
19	207	11.1	1959	12	ADG22400 Human sof
20	206.8	11.1	1307	3	AAAG63927 DNA encod
21	148	8.0	148	10	AAL51522 Human Nkd

22	137	7.4	598	12	ACH74795 Human gen
23	137	7.4	1950	6	ABG73820 Human col
24	134	7.2	134	10	AAL51518 Human Nkd
25	133	7.2	133	12	ACH88495 Human gen
26	128	6.9	128	10	AAL51524 Human Nkd
27	120	6.5	120	10	AAL51514 Human Nkd
28	109	5.9	639	8	ACD05637 cDNA enco
29	107	5.8	107	10	AAL51520 Human Nkd
30	96	5.2	96	10	AAL51521 Human Nkd
31	89	4.8	554	12	ACH74420 Human gen
32	86.4	4.6	179	12	ACH88120 Human gen
33	85	4.6	85	10	AAL51523 Human Nkd
34	72.4	3.9	1817	6	ABK34945 Human cDN
35	70.2	3.8	2723	4	AAK75944 Human imm
36	68.4	3.7	1264	12	ADG24765 Dog TRAIL
37	67	3.6	67	10	AAL51519 Human Nkd
38	65.6	3.5	670	6	ABV73487 Human cyt
39	62.8	3.4	551	10	ADFG1081 Leukaemia
40	60.6	3.3	2000	8	ADA71938 Rice gene
41	58.8	3.2	707	6	ABX04893 Cornu BP
42	58.2	3.1	1163	12	ADG24767 Cat TRAIL
43	57.4	3.1	417	6	ABV73485 Human cyt
44	57.2	3.1	417	6	ABV73496 Human cyt
45	56.8	3.1	114955	2	AAK53491 Human ade

ALIGNMENTS

RESULT 1	
AAL51511	
ID AAL51511 standard; DNA; 1859 BP.	
XX	
AC AAL51511;	
XX	
DT 24-APR-2003 (first entry)	
XX	
DE Human Nkd (nmkd) protein coding sequence.	
XX	
KW Human; gene; ds; gene therapy; Nkd; nmkd; wnt signalling pathway; cancer;	
KM colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer.	
XX	
OS Homo sapiens.	
XX	
PH Key	Location/Qualifiers
FT CDS	224..1636
FT	/*tag= a
FT	/product= "Human Nkd protein"
XX	
PN W0200292832-A2.	
XX	
PD 21-NOV-2002.	
XX	
PE 27-NOV-2001; 2001WO-US044092.	
XX	
PR 27-NOV-2000; 2000US-0252884P.	
PR 16-MAY-2001; 2001US-0291109P.	
XX	
PR 01-OCT-2001; 2001US-0325571P.	
XX	
PA (CHIR) CHIRON CORP.	
XX	
PI Rohan M, Chan V, Yan D,	
XX	
DR WPI; 2003-129303/12.	
XX	
DR P-PSDB; AAO16389.	
XX	
PT New human and non-human primate homologues of Nkd protein, and Nkd genes,	
PT useful for treating cancer involving aberrant wnt signalling, e.g. colon	
PT cancer, head and neck cancer, ovarian cancer, or breast cancer.	
XX	
PS Claim 4; Fig 3; 99gp; English.	
XX	
CC The invention comprises the amino acid and coding sequence of the human	

CC Nkd protein (hNkd) - a regulator of Wnt signalling pathways. The hNkd DNA
CC and protein sequences are useful for the treatment of cancer involving
CC aberrant Wnt signalling (e.g. colon cancer, head and neck cancer, ovarian
CC cancer and breast cancer). The present DNA sequence encodes the human Nkd
CC protein of the invention

XX Sequence 1859 BP; 434 A; 587 C; 551 G; 287 T; 0 U; 0 other;

Query Match 100.0%; Score 1859; DB 10; Length 1859;

Best Local Similarity 100.0%; Freq. No. 0;

Matches 1859; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAATTCGCGCTTCAATACGACTCACTATAGGCGAAGCATGTATACAAACGACAGATACG 60
Db 1 GAATTCGCGCTTCAATACGACTCACTATAGGCGAAGCATGTATACAAACGACAGATACG 60
Qy 61 CGGGAGCTCGGCG 120
Db 61 CGGGAGCTCGGCG 120
Qy 121 CTGGGCTCGGGGCGCTGCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
Db 121 CTGGGCTCGGGGCGCTGCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
Qy 181 GGGCGCATGCTTAGGAGACGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 181 GGGCGCATGCTTAGGAGACGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Qy 241 CAAGCCG 300
Db 241 CAAGCCG 300
Qy 301 CGCTGCGTGGGCG 360
Db 301 CGCTGCGTGGGCG 360
Qy 361 TGTCTCGGAGACCCCGAGAGCTGCGGTTGACGAGACCATATAGCCGGAAGCACCCGGAG 420
Db 361 TGTCTCGGAGACCCCGAGAGCTGCGGTTGACGAGACCATATAGCCGGAAGCACCCGGAG 420
Qy 421 CGTGGCGAGCGTGTGAGAGACAGCTCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Db 421 CGTGGCGAGCGTGTGAGAGACAGCTCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Qy 481 AGTGGCGCGTGTGAG 540
Db 481 AGTGGCGCGTGTGAG 540
Qy 541 GAGAGTAGCGAAGCTTGCAGAGCTCCAGAGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Db 541 GAGAGTAGCGAAGCTTGCAGAGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Qy 601 CGAGCTGTCAATGAG 660
Db 601 CGAGCTGTCAATGAG 660
Qy 661 CAACAAAGGAG 720
Db 661 CAACAAAGGAG 720
Qy 721 GGTGAGCT 780
Db 721 GGTGAGCT 780
Qy 781 CGTGGCGCGAG 840
Db 781 CGTGGCGCGAG 840
Qy 841 GAGCGCAAG 900
Db 841 GAGCGCAAG 900
Qy 901 GCAAGGAG 960

Db 901 GCAAGGAG 960
Qy 961 CCAATTCGCTGATGAG 1020
Db 961 CCAATTCGCTGATGAG 1020
Qy 1021 AAACCTACAGCTCCCAATTTGGGAGCTGAGCTCCCTCCGAGAGAGAGAGAGAGAGAGAG 1080
Db 1021 AAACCTACAGCTCCCAATTTGGGAGCTGAGCTCCCTCCGAGAGAGAGAGAGAGAGAGAG 1080
Qy 1081 CCCCAGACCTCCCAATTCGAGTCTGCTCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Db 1081 CCCCAGACCTCCCAATTCGAGTCTGCTCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Qy 1141 ACAACGAG 1200
Db 1141 ACAACGAG 1200
Qy 1201 CAAGGTCTCAGAGCTCCAG 1260
Db 1201 CAAGGTCTCAGAGCTCCAG 1260
Qy 1261 GAGGTCCCGCAAG 1320
Db 1261 GAGGTCCCGCAAG 1320
Qy 1321 AAACAG 1380
Db 1321 AAACAG 1380
Qy 1381 CAGCGCGAG 1440
Db 1381 CAGCGCGAG 1440
Qy 1441 CAAG 1500
Db 1441 CAAG 1500
Qy 1501 CTTGGGAG 1560
Db 1501 CTTGGGAG 1560
Qy 1561 GAGCGCGTCAAG 1620
Db 1561 GAGCGCGTCAAG 1620
Qy 1621 CTTCTACAG 1680
Db 1621 CTTCTACAG 1680
Qy 1681 CCGGACACCAAG 1740
Db 1681 CCGGACACCAAG 1740
Qy 1741 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1800
Db 1741 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1800
Qy 1801 CACGAACTAACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1859
Db 1801 CACGAACTAACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1859

RESULT 2
AAA63925
ID AAA63925 standard; DNA; 1438 BP.
XX
AC AAA63925;
XX
XX 04-DEC-2000 (first entry)
XX
XX DNA encoding protein related to Drosophila naked cuticle polypeptide.

XX Nkd gene; naked cuticle gene; segment-polarity gene; Wnt signalling;
KW gene therapy; Nkd defect; cancer; ss.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 8..1420
FT /tag= a
FT /product= "naked cuticle polypeptide"
XX
PN MO200049034-A1.
XX
XX 24-AUG-2000.
XX
XX 17-FEB-2000; 2000MO-US004188.
XX
XX 17-FEB-1999; 99US-0120646P.
XX
XX (STRD) UNIV LEIAND STANFORD JUNIOR.
XX
XX Scott M, Zeng W, Wharton K;
XX
XX MPI: 2000-571967/53.
XX P-PSDB; AAB08216.
XX
XX An isolated nucleic acid molecule useful for analyzing (genetic
PT predilection to) a disease state and for therapeutic purposes e.g.
PT treatment of cancer comprises a sequence encoding a naked cuticle
PT protein.
XX
XX Claim 3: Page 46-48; 58pp; English.
XX
XX The present sequence encodes a protein related to the Drosophila Nkd
CC (naked cuticle) polypeptide. In Drosophila, Nkd is a segment-polarity
CC gene whose expression is induced by Wnt signalling. The Nkd polypeptide
CC acts to antagonize Wnt signalling. Nkd may link ion fluxes to the
CC regulation of Wnt signal potency, duration or distribution. The Nkd
CC polynucleotide can be used for identifying homologous or related
CC proteins, to modulate the expression or function of Nkd polypeptides, and
CC in studying associated physiological pathways. Nkd polynucleotides can
CC also be used in gene therapy to treat disorders associated with Nkd
CC defects. They may also be used for therapeutic purposes e.g. treatment of
CC cancer
XX
SQ Sequence 1438 BP; 319 A; 478 C; 439 G; 202 T; 0 U; 0 Other;
Query Match 77.2%; Score 1434.8; DB 3; Length 1438;
Best Local Similarity 99.9%; Pred. No. 3.9e-277;
Matches 1436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 217 CCCACGATGGGAACTTCACTCAAGCGCGCGCTGTGCAAGCGAGGAGAGCCC 276
DB 1 CCCACGATGGGAACTTCACTCAAGCGCGCGCTGTGCAAGCGAGGAGAGCCC 60
QY 277 GGAAGGTGACAGCTTGCCTGAGCGCTGCTGGGACTTGGAGGGCATGAGAGTGGAT 336
DB 61 GGAAGGTGACAGCTTGCCTGAGCGCTGCTGGGACTTGGAGGGCATGAGAGTGGAT 120
QY 337 CGGGAACAGCGCTGCGCGCGCGCTGTGCGGACCCCGACAGCTGGGTGGCGGCGAC 396
DB 121 CGGGAACAGCGCTGCGCGCGCGCTGTGCGGACCCCGACAGCTGGGTGGCGGCGAC 180
QY 397 CATAGCCGGAAGACCCGAGAGCTGCTGGGCGAGCTGTGAGAGACAGCTCAGCGAGGA 456
DB 181 CATAGCCGGAAGACCCGAGAGCTGCTGGGCGAGCTGTGAGAGACAGCTCAGCGAGGA 240
QY 457 AGAGAGAGAGCATTTTGGCTGGAAGTGGCCCTGCTCTTGAGAAAGTGAACGGGCTGGG 516
DB 241 AGAGAGAGAGCATTTTGGCTGGAAGTGGCCCTGCTCTTGAGAAAGTGAACGGGCTGGG 300
QY 517 CACCGAGATGAGAAAGATGAGAGAGTGAAGGAACTTGGCCAGGCTCCAAGAGCA 576
|||||

DB 301 CACCGAGATGAGAAAGATGAGAGAGTGAAGGAACTTGGCCAGGCTCCAAGAGCA 360
QY 577 GCTGAAGTTGAAAGAGCTCCAGTGCAGCTGTCCATGAGAGAGACAGCCGGGAGAGTG 636
DB 361 GCTGAAGTTGAAAGAGCTCCAGTGCAGCTGTCCATGAGAGAGACAGCCGGGAGAGTG 420
QY 637 GACCTTCACCCCTGATGACTTTGACAAACAGGCAAGGTCAACCCGAGAGACATCAGG 696
DB 421 GACCTTCACCCCTGATGACTTTGACAAACAGGCAAGGTCAACCCGAGAGACATCAGG 480
QY 697 CTGCTGACACCATCTATGAGGTGTGAGCTCTCTGTCAACCACTCCCAATCAG 756
DB 481 CTGCTGACACCATCTATGAGGTGTGAGCTCTCTGTCAACCACTCCCAATCAG 540
QY 757 CAAGATGCTGGGATTAAGCTCAACCTGACCCCGATGAGGCAAGAGAGAGAGCT 816
DB 541 CAAGATGCTGGGATTAAGCTCAACCTGACCCCGATGAGGCAAGAGAGAGAGCT 600
QY 817 CTTTGTCAATCAAGCTGACCTGAGAGCGGCAAGGCCCCGAGAGAGCAAGGCCACTGA 876
DB 601 CTTTGTCAATCAAGCTGACCTGAGAGCGGCAAGGCCCCGAGAGAGCAAGGCCACTGA 660
QY 877 GAACTGCGGAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 936
DB 661 GAACTGCGGAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 937 CTTGAGAGAGCTTGGCTGCTACCAACCATGCTGATGAGAGATGAGAGAGAGAGAG 996
DB 721 CTTGAGAGAGCTTGGCTGCTACCAACCATGCTGATGAGAGATGAGAGAGAGAGAG 780
QY 997 CTACTTGAATCTGCGCGGATGAGAAACATCAAGCTCCAAATTTGGGCTGCTCCCTTC 1056
DB 781 CTACTTGAATCTGCGCGGATGAGAAACATCAAGCTCCAAATTTGGGCTGCTCCCTTC 840
QY 1057 CGAGGCCAGAGATCAAG 1116
DB 841 CGAGGCCAGAGATCAAG 900
QY 1117 TGAAGCCGAAAGCCATCCATCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1176
DB 901 TGAAGCCGAAAGCCATCCATCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
QY 1177 CCACTTCTTGAAG 1236
DB 961 CCACTTCTTGAAG 1020
QY 1237 CCAAG 1296
DB 1021 TCAAG 1080
QY 1297 GAGGCAAGTGGCCAG 1356
DB 1081 GAGGCAAGTGGCCAG 1140
QY 1357 GTCCCTCTGCGCCCACTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1416
DB 1141 GTCCCTCTGCGCCCACTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1417 GCAAG 1476
DB 1201 GCAAG 1260
QY 1477 ACCAGTGGCTCAAGTGGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1536
DB 1261 ACCAGTGGCTCAAGTGGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 1537 GGTGTGTATGAG 1596
DB 1321 GGTGTGTATGAG 1380
QY 1597 TGAACATCAACCACTTACCACTTACCACTTACCACTTACCACTTACCACTTACCA 1654
DB 1381 TGAACATCAACCACTTACCACTTACCACTTACCACTTACCACTTACCACTTACCA 1438
|||||

RESULT 3
AAA63924
ID AAA63924 standard; DNA; 1731 BP.
XX AC
XX AAAG3924;
XX DT 04-DEC-2000 (first entry)
XX DE DNA encoding protein related to Drosophila naked cuticle polypeptide.
XX Nkd gene; naked cuticle gene; segment-polarity gene; Wnt signalling;
XX Kwd therapy; Nkd defect; cancer; ss.
XX Mus musculus.
OS
XX Key Location/Qualifiers
FT CDS 140..1555
FT FT /*tag= a
FT FT /product= "naked cuticle polypeptide"
PX WO20049034-A1.
PX PD 24-AUG-2000.
PX PF 17-FEB-2000; 2000WO-US004188.
PX PR 17-FEB-1999; 99US-0120646P.
PX XX (STRD) UNIV LELAND STANFORD JUNIOR.
PA Scott M, Zeng W, Wharton K;
PX MPI; 2000-571967/53.
DR P-PSDB; AAB08215.
XX An isolated nucleic acid molecule useful for analyzing (genetic
PT predisposition to) a disease state and for therapeutic purposes e.g.
PT treatment of cancer comprises a sequence encoding a naked cuticle
PT protein.
XX
PS Claim 3; Page 43-45; 58pp; English.
XX
CC The present sequence encodes a protein related to the Drosophila Nkd
CC (naked cuticle) polypeptide. In Drosophila, Nkd is a segment-polarity
CC gene whose expression is induced by Wnt signaling. The Nkd polypeptide
CC acts to antagonize Wnt signaling. Nkd may link ion fluxes to the
CC regulation of Wnt signal potency, duration or distribution. The Nkd
CC polynucleotides can be used for identifying homologous or related
CC proteins, to modulate the expression or function of Nkd polypeptides, and
CC in studying associated physiological pathways. Nkd polynucleotides can
CC also be used in gene therapy to treat disorders associated with Nkd
CC defects. They may also be used for therapeutic purposes e.g. treatment of
CC cancer
XX
SQ Sequence 1731 BP; 414 A; 521 C; 504 G; 292 T; 0 U; 0 Other;
Query Match 65.7%; Score 1221.4; DB 3; Length 1731;
Best Local Similarity 83.9%; Pred. No. 2e-234;
Matches 1460; Conservative 0; Mismatches 256; Indels 25; Gaps
OY 82 CGGCGCAGAGAGCGGTCCGGCCGCCCTTCGGGCTTCCGGCTTGGGGGCTGCTTG 141
Db 6 CCGCGCCCGAGCGGTCCGGCCGCCCTTAAGGCTGCGCTCGGCGCGACTG--TGA 63
OY 142 GGAGGAGAGAGCAAGAAGAGCGCACGACGGCCGCGGCGCATGGCTTAGAGAGC 201
Db 64 GGAGGAGCGAGCGA--GGCTGCGCCGGGCTCGGCGCGCGACGATGGCTTAGAGCGCT 121
OY 202 TTCGCGCGCGCGACGCCCGAGCATGAGGAAATTCACTTCAAAGCCGCGCGTGTGCAA 261
Db 122 CCGCGCGCGCGC---GCCCGAGCATGAGGAAATTCACTTGAAGCCGCGCGCGTGTGCAA 177

QY	262	GCGCAGGGAGAACCCCGAAAGGTGACAGCTTGCCCGTGAAGCGCTGCTGGGCTCGGAAGG	321
Db	178	GCGCAGGGAGAGCCCGGAAGGTGACAGCTTTCGTGAAGCGCTGCTTGGGCAAGAAAGG	237
QY	322	CATGAGAGAGTGGATTCGGGAGACAGCGCGCCCGGCGGCTGTCTGGAGCCCGACAGCT	381
Db	238	CATCGAGAGTGGATTCGGGAGGCGCGCTGTCTAGGCGAGCGTCTCAGAGCCCCGTCAGCT	297
QY	382	GCGGTTGGCGGGCACCATAGGCCGAAGCACCCGGAGCTGTGTGGCGACGTGTGTAGAGA	441
Db	298	GAGATTTGGCAAGCACGTGTGTGGCGACGCTCGGGAACTCGTGGGGTGACACTTCTTAGAGA	357
QY	442	CACGCTCACCGAGGAAGAGAGACGACTTTCGCGCTGGAAGTGGCCCTGCTCTTAGAGA	501
Db	358	GGCTTCGTGTAGAGGAGGACGAGAGACGACTTCCCTTAAGATGGCCCTGCGCGCTTAGAGA	417
QY	502	GACTGACGGGCTGGGCGACCGAGATGGAAGAAGATGGAAGAGTGAACGGAACCCCTGGCC	561
Db	418	GATCGACAGCCTTAGGTAGTGGAGATGAAGAAGATGGAAGAGACTGACGGAACCTGGCCA	477
QY	562	AGGCTCCAGAAGCAGCTGAAGTTTGAAGAGCTCCAGTGCACGCTGTCAATGAGAGAGA	621
Db	478	GGCTTCAGAAGAGCAGCTCAAGTTTGAAGAGCTAAGTGTGATGTCTGTGTGAAGAGGA	537
QY	622	CAGCCGGCAGAGTGGACCTTTCACCTGTATGACTTTTGAACAACAACGCGAAGTCAACCG	681
Db	538	CAGCCGGCAAGTGGACTTTTCACTCTATATGACTTCCACAACAATGGCAAGTATACCG	597
QY	682	AGAGGACATCACAGACTTGTGTACACCATCTATAGAGTGTGTGACTCTCTGTCAACA	741
Db	598	TGAGGACATTTACCACTTGTGTGTGCATATCATATGATGTGTGACTCTCTGTCAACA	657
QY	742	CTCCCAACATATCCAGACAGATGCTGCGGGTAAAGCTCACCGTGCGCCCGATGCGAGCA	801
Db	658	TTCCCCACATATCAGACAGACATCTGCGGGTAAAGTCAACGTGGCTCTGTACGGGAGCA	717
QY	802	GAGCAAGAGAGCGTCTTGTCAATCAGCTTGACCTGCGAGACGCAAGGCCCGCGAGAGA	861
Db	718	GAGTAAAGAGAGCGTCTTTTCAACCATACGATCTGCAGAGCACAAAGCCCCGAGAGAGA	777
QY	862	GACCAAGCCCACTGAGGACTGTGCGGAGCTGGGAGAAAGACAGCGAGCCCGCTCAGGTT	921
Db	778	CACCAAAACCCCGCTGAGGAGCTGCGGTGGGAGAAAGACGCGAGGCCCACTCAGGTT	837
QY	922	CCAGGGGTACAAGCGCGCTGAGGAGACTGTGGGCTGCTACACAACAATTGCGTATGAGAACAT	981
Db	838	CCAGGGTACAAGCACTGAGGAGCGCAAGACTGTCTACACAACAATTGCGTATGAGAACAT	897
QY	982	CGAGAGAGAAACCACTACTTAGATCTGTGCGCGGGATAGAAAACTACAGTGTCCCAATTGG	1041
Db	898	TGAGAGAGAAACCACTACTTAGACTGTGCGGGGATAGAAACTACAGTGTCAATTGG	957
QY	1042	GCTGGGCTCCCTTCCGTGGGCCAGAAAGTCAGAACGTGCCCCCGGCACTTCCATCCAC	1101
Db	958	ACCGGAGATCCCTTGGTGTGGCCAGAAAGTCAAGAGCTGCCCTCGAATCTCCAAACCCAC	1011
QY	1102	TGATATCTGCTCCATATGAGCCGGAAGCCATCTCAATCCCAACGAAAGCCCAAGGGGT	1161
Db	1018	TGCTCTTGTGCTCCACGAGCAGAAAGCTGCCCACTATCCCAACCGAAGGCCCAAGGGGT	1077
QY	1162	GGACCCGGGCTCTCTTCACTTCTTGAACACCCCAATGCGCAAGGTCCTCAGAGCTCCAGCA	1221
Db	1078	GGACCCAGGCTCTCTTCACTCTTGAACACCCCAATTTGCGCAAGGTCATCAGAGCTCCAGCA	1133
QY	1222	AACGCTCCGGGAGCACCCAGAACGGGAGCAACCTTTGTGAAGTCCCCCAAGGCCAGGG	1281
Db	1138	AACGCTCCGGGAGCACTCAGAGATGGGAGCAAGCACTTTGTGAAGTCCCCCAAGGCCAGGG	1199
QY	1282	CAGAGATGTGGTGTGGGCAAGTGTGGCCAGAGGGGCAAGAAACAAGCCCCCTCTGTGGACC	1341
Db	1198	CAAGAAATGTGGTATGTGGGCAAGGGGCCAGAGGTGCAAGAACAGGCTTCACTGTATGCC	1251

QY	1342	CGCCATCCCTGGGTGTGTCCCTTCGACCACCTGGCTCCAGCCCGCTCTCTCCCTC	1401
Db	1258	CACCACCCATCTGTCTCCCTCTGCCCCTGAGCCAGCCCGCTCTCTCCCTC <td>1317</td>	1317
QY	1402	CCTGAGCCCCCTCGGGGACAGAAGACACAGACCCAGGACAGAGGAGCCAGAGGGCTG <td>1461</td>	1461
Db	1318	CTTGACACCTCTGGGGACAGAGAAACACAGAGCATGAGCCAGAGGAGCCAGGCGAGCTG <td>1377</td>	1377
QY	1462	CCGGGGGCTGAGGACCACTGTGGCTCAGTGG--CCCTGTCTCGGGGCGGAGCACT <td>1518</td>	1518
Db	1378	CCGGGGGCTGAGGGCCCCCTGGCTGAGAGGGCTCCACCGTCATGGGGCGGAGCAGGT <td>1437</td>	1437
QY	1519	GCGGAGCTGCCCCCTGTTGGTGTGTATGAGAGCCAGGCGGAGCCGCTCCAGAGCA <td>1578</td>	1578
Db	1438	GAGGAGCTGCTGCTGCGGTGATGTGTATGAGAGCCAGGCTAGGCGGCTCCAGAGACA <td>1497</td>	1497
QY	1579	TGAGGACACAGACCAACCATGAAATTCACCATTAACAACAATTCTTCCAGACATAGAG <td>1638</td>	1638
Db	1498	CGAAGACCATATACACACAGAAACATCACACCATTTATTCACACTTCTATCCAGCCCTAGAC <td>1557</td>	1557
QY	1639	CCCCCTCCCAAGGGCCCCACCCCTGECATATGAAAGACCCCACCCCGACACAGAAAGCAT <td>1698</td>	1698
Db	1558	CC-----CAGCAGGCTGCCACGGGAAAGAACCCAGCC--ACACCTTAAGGCAT <td>1603</td>	1603
QY	1699	TATTATTCTAATTAATTAATTTGTTATTTATGATGATTAATTTGTTAATAATAATTAATTTACTC <td>1758</td>	1758
Db	1604	TATTATTCTAATTAATTAATTTGTTATTTATGCAATTAATTTGTTAATAATAATTAATTTACTC <td>1663</td>	1663
QY	1759	CACCTAATATTAGCTAGGCTACATGTAGAAATCTATGTGAAACACAGAACTTAACCTTTTA <td>1818</td>	1818
Db	1664	CACCTAATATTAGCTAGGCTACATGTAGAAAGACATGTGAAACACAGAACTTAACCTTTTA <td>1723</td>	1723
QY	1819	T 1819 <td></td>	
Db	1724	T 1724 <td></td>	
RESULT 4			
AAH28343			
ID	AAH28343 standard, DNA; 1416 BP.		
AC	AAH28343;		
XX			
DT	05-SEP-2001 (first entry)		
XX			
DE	Nucleotide sequence of a murine mnkd protein.		
XX			
KM	Dishevelled-associated protein 1A; DAPIA; mnkd; dishevelled protein;		
KW	Wnt signalling; colon cancer; cancer; ss.		
XX			
OS	Mus sp.		
XX			
FN	Key		
FT	CDS		
FT	Location/Qualifiers		
FT	1..1416		
FT	/*tag= a		
FT	/product= "mnkd"		
XX			
PN	W0200144279-A2.		
XX			
XX	21-JUN-2001.		
PD			
PF	05-DEC-2000; 2000MO-US032986.		
XX			
PR	17-DEC-1999; 99US-0172434P.		
XX			
XX	(CHIR) CHIRON CORP.		
PA			
PI	Yan D, Williams LT;		
XX			
DR	WPI, 2001-425440/45.		
DR	P-PSDB; AAB84645.		
XX			
DT	Novel mammalian dishevelled-associated proteins, mnkd and DAPIA, useful		

PT		for inhibiting Wnt signaling in mammalian cells, and thus for treating colon cancer.
XX		Claim 2; Fig 1; 67pp; English.
CC		The present sequence encodes a murine mKd protein. The specification describes mKd and dishevelled-associated protein 1A (DAP1A) proteins, that interact with mammalian dishevelled protein. mKd is useful for inhibiting Wnt signaling in mammalian cells, and thus for treating colon cancer. mKd is also useful for activating the JNK pathway. mKd and DAP1A are also useful for screening drugs that are useful for treating cancer.
CC		
XX		
SQ	Sequence 1416 BP; 339 A; 434 C; 419 G; 224 T; 0 U; 0 Other;	
Query Match	58.0%; Score 1078.4; DB 4; Length 1416;	
Best Local Similarity	85.6%; Pred. No.7.3e-206;	
Matches 1212; Conservative	0; Mismatches 201; Indels 3; Gaps 1	
OY	224 ATGGGAAATCTTCACTCCAGCCCGGCCCGCCTGTGTCAAGCCGACAGAGAGCCCGGAAGT	283
DB	1 ATGGGAAATCTTCACTCGAAGCCCGGCCCGCCTGTGTCAAGCCGACAGAGAGCCCGGAAGT	60
OY	284 GACAGCTTCGCCTGTAGCGCTTGCCCTGCGCTCGAAGGSCATCGAGAAGTGAATCGGAGAG	343
DB	61 GACAGCTTTGCTGTAAAGCGCTGTCTTGGGCAAGAAAAGCATCGAGAAGTGAATCGGAGAG	120
OY	344 CAGGCTGCCCCGGGCGGTGTCTCGGAGACCAGACGTCGCGTGTGGCGGGACCAATAAGC	403
DB	121 CAGGCTGTCCAGGAGAGCGTCTCAAGAACCCCGTCAAGCTGAATTGGCAGGCACTGTGGT	180
OY	404 CGAAGCACCCGAGAGACTGTGTGAGGAGACGCTGTGAGAGACAAGCTCAGCGAGAAAGAGAG	463
DB	181 CGAGGCAATCGGGAACTGTGTGGTAGACCTTCTAGAGAGCGTCTCGTAGAGAGAGACGAG	240
OY	464 GACGACTTTGCGAGTGAAGTGGCCCTGCTCTTGAAGAACTGAACGCGCTGGGACGCGGA	523
DB	241 GACGACTTCCCCCTAGAAAGTGGCCCTGCGCTGAGAAAGATCGACAGCTGTAGTAGTA	300
OY	524 GATGAGAAAGATGAGAGAGAGTAGAGAGAACCTTCGCCAGGCTCCAAGACAGCTGTAAG	583
DB	301 GATGAGAAAGAAATGAGAGAGACTAGAGGCAACTGTGGCCAGGCTCCMAAMACAGCTCTAAG	360
OY	584 TTGGAAGAGCTCCAGTGTGCACTGTGCAATGAGAGAGACACCGGAGAGAGTGAACCTTC	643
DB	361 TTGGAAGAGCTACAGTGTGATGTCTTGTGTGAGAGAGACACCGGAGAGAGTGAACCTTC	420
OY	644 ACCCTGTATGACTTTGACACMACAGGAGTCAACCGGAGAGACATCACAGCTTGTCTG	703
DB	421 ACTGTATATGACTTTCAGACACAAATGTGGAAGTGAACCGTGTAGAGACATTACAGCTTGTG	480
OY	704 CACACCATCTATGAGTGTGTGAATCCTCTGTCAAACAATCCCAATCATCAGCAAGATG	763
DB	481 CATACCATCTATGAAAGTGTGTGAATCCTCTGTGTGAACAATCCCACATCAAGCAAGACA	540
OY	764 CTGGGGGGAAGCTCAACCGTGGCCCCGATGTGGACCAAGACAGAGAGAGGTCCTGTGTC	823
DB	541 CTGGGGGGAAGCTCAACCGTGGCTCTTGAACGAGAGCCAGATTAAGAGAGGTCCTTTTC	600
OY	824 AATCAGGCTGACTGTGAGAGCGCAAGGCCCGGACAGAGACCAAGCCCATGTAGAGACTTG	883
DB	601 AACCATACCAATCTGACAGAGACCAAGGCCCGGACACACCAAAACCCGCTGAGAGAGCTG	660
OY	884 CGAGCTGTGGAGAAAGACAGAGAGCCCGCTCAGGTTCCAGGGTGACAGCCGCTCGAG	943
DB	661 CGTGGCTGTGGAGAAAGACAGCGAGCCCACTCAGGTTCCAGGGTGACAGCCACTCGAG	720
OY	944 CAGTCTGGGCTGTACCAACATTTGGTGTAGTGAAGAACATCGAGAGAGAAACCACTACTTA	1003
DB	721 CAGCCTAGCTGTACCAACCATTTGGTGTAGTGAAGAACATTGAGAGAGAAACCACTACTTA	780
OY	1004 GATCTCGCCGGGATAGAAAACTACAGCTCCCAATTTTGGGCTGTGCTCCCTTCGTGGCC	1063

```

Db      781 GACCTGGCGGGATAGAGAACTACAGCTCAAGTTTGAACCGGATCCCTTCGGTGGCC 840
QY      1064 CAGAAATGCAAACTGCCCCCGGCACTCCAAATCCCATGATCTGGCTCCATGAGCCG 1123
Db      841 CAAAGTCAGAGCTGCCCCCTCGAATTCACACCCCACTGCTCTGCTCCCAAGAGCCA 900
QY      1124 GAAGCCATCCACATCCCAACCGAAAGCCCAAGGCGGTGAGACCGGCTCTTCACTTC 1183
Db      901 GAAGCTGCCCACTCCACACACCGAGGCCCAAGGTGAGACCAAGGCTCTTCACTTC 960
QY      1184 CTTGACACCCCAATCGCCAAAGTCTCAGAGCTCCAGAACGGCTCCGGGGCAGCCAGAGC 1243
Db      961 CTTGACACCCCAATTCGCAAGGCTCAGAGCTCCAGAACGGCTCCGGGGCAGCTCAGAGT 1020
QY      1244 GGGAGCAGACATTTTGTAGAGTCCCCCAAGGCCCAAGGCAAGATGTGGGTGTGGGCAC 1303
Db      1021 GGGAGCAGACATTTTGTAGAGTCCCCCAAGGCCCAAGGCAAGATGTGGGTGTGGGCAC 1080
QY      1304 GTGGCCAGAGGGGCAAGAAACAAAGCCCTCTGAGGACCCGCAATCCCTGCGGTGTCGCC 1363
Db      1081 GGGGCCAGAGGTGCAAGAACGAGCTCCACTGGTACCCACCACTACTGTCTCTCCGCC 1140
QY      1364 TCGGCCCACTGGCTGCGCAGCCCGGCTCTCTCCCTCCCTAGCCGCCCTCGGGCAGAG 1423
Db      1141 TCGGCCCATCTGGCCCAAGCCAGCCCTTCTCCCACTGGGCACTCCCTGGGGCAGAG 1200
QY      1424 AAGCACAAGACCGAGAGCAGAGAGCAGCAGAGGCTGCGGGGCTTGCAGGCACTG 1483
Db      1201 AAGCACAAGACCGAGAGCAGAGAGCAGAGCAGAGGCTGCGGGGCTTGCAGGAGCCCTG 1260
QY      1484 GCCTCAGAGTGA---CCCTGTCTGGGGGAGAGACCTGCGGAGCTGCGGCTTGTG 1540
Db      1261 GCTGCGAGGAGGCTCCACCTGTATGAGGCGGGAGAGGTGAGGAGCTGCTGCGGTGTG 1320
QY      1541 GTGTATGAGAGCCAGCGCGGCGAGCCGCTCCAGAGATGAGACATGACCAACCATGAA 1600
Db      1321 GTGTATGAGAGCCAGCGCTGGGCGAGCGCTCCAGAGACAGCAACCATGCAAGAA 1380
QY      1601 CATCACCAACATTAACCACTTCTTACAGACATAG 1636
Db      1381 CATCACCAACATTAACCACTTCTTACAGACCTAG 1416

```

RESULT 5

AAL51512

ID AAL51512 standard; DNA, 1401 BP.

XX

AC AAL51512;

XX

DT 24-APR-2003 (first entry)

XX

DE Mouse Nkd protein gene sequence.

XX

XX Mouse; gene; ds; gene therapy; Nkd; hnkd; Wnt signalling pathway; cancer;

XX

XX colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer;

XX

XX murine.

XX

XX Mus SP.

XX

XX WO200292832-A2.

XX

XX 21-NOV-2002.

XX

XX 27-NOV-2001; 2001WO-US044092.

XX

XX 27-NOV-2000; 2000US-0252884P.

XX

XX 16-MAY-2001; 2001US-0291109P.

XX

XX 01-OCT-2001; 2001US-0325571P.

XX

XX (CHIR) CHIRON CORP.

XX

XX Rohan M, Chan V, Yan D;

XX

```

DR      WPI; 2003-129303/12.
XX      New human and non-human primate homologues of Nkd protein, and Nkd genes,
PT      useful for treating cancer involving aberrant Wnt signaling, e.g. colon
PT      cancer, head and neck cancer, ovarian cancer, or breast cancer.
XX      Example 2; Fig 4; 99p; English.
XX      The invention comprises the amino acid and coding sequence of the human
CC      Nkd protein (hnkd) - a regulator of Wnt signaling pathways. The hnkd DNA
CC      and protein sequences are useful for the treatment of cancer involving
CC      aberrant Wnt signaling (e.g. colon cancer, head and neck cancer, ovarian
CC      cancer and breast cancer). The present DNA sequence represents the mouse
CC      Nkd gene
XX      Sequence 1401 BP; 336 A; 430 C; 416 G; 219 T; 0 U; 0 Other;
SQ

```

Query Match 57.4%; Score 1066.6; DB 10; Length 1401;

Best Local Similarity 85.6%; Pred. No. 1.7e-203;

Matches 1199; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

```

QY      224 ATGGGGAATCTTCACTCCAAAGCCGCGCTGTGCAAGGCAAGGAGAGCCGGAAAGT 283
Db      1 ATGGGGAATCTTCACTCCAAAGCCGCGCTGTGCAAGGCAAGGAGAGCCGGAAAGT 60
QY      284 GACAGCTTGGCGGTGAGCGCTGCTGCGGAGCTCGGAAAGGCAATCGAGAGTGGAGAG 343
Db      61 GACAGCTTGTCTGTAAAGCGCTGCTTGGGCAAGAAAGCATCGAGAGTGGAGAG 120
QY      344 CAGCGCTGCGCGGCGGTGTCTCGGAGACCCGCAAGCTGCGGTGGCGGACCAATAGGC 403
Db      121 CAGCGCTGTCAGAGCAGCGCTCTCAGAGACCCGCTCAGCTGAGATTGGCAGGCACTGTGT 180
QY      404 CGAAGCACCCGGAGAGCTGCTGGGCGAGCGTGTGAGAGACAGCGTCCAGGAGAGAGAG 463
Db      181 CGAAGCACCTGGGAACTGTGGGAGACATTTGAGAGGCTCTCGTGGAGAGAGAG 240
QY      464 GACGACTTTGGCTGGAAGTGGCGCTGCTCTGAGAGACTGACCGGCTGGGAGCGGA 523
Db      241 GACGACTTCCCTTGAAGAGTGGCGCTGCTGAGAGAGTGGCGCTGAGAGAGTGGAG 300
QY      524 GATGAGAAAGATGAGAGAGTGAAGCAACCTGCGGAGCTTCCAAAGAGAGTGAAG 583
Db      301 GATGAGAAAGATGAGAGAGTGAAGCACTGGCCAGGCTTCCAAAGAGAGTGAAG 360
QY      584 TTGGAAGAGCTCCAGTCCGAGCTGTCCATGAGAGAGAGCCGGGAGAGTGGAGCTTC 643
Db      361 TTGGAAGAGCTCAAGTGTGATGCTCTGTGAGAGAGAGAGCCGGGAGAGTGGAGCTTTC 420
QY      644 ACCGTGATGACTTTGACAAACAGGAGAGTCAACCCGAGAGAGCATCAACAGCTTGTG 703
Db      421 ACTGTATATGACTTTGACAAACAGGAGAGTCAACCCGAGAGAGCATCAACAGCTTGTG 480
QY      704 CACACCATATGAGAGTGTGAGCTCTCTGTCAACCATCTCCCAATCCAGCAAGATG 763
Db      481 CATACCATATGAGAGTGTGAGCTCTCTGTCAACCATCTCCCAATCCAGCAAGATG 540
QY      764 CTGGGGGTAAAGCTCAACCGTGGGCGCGGAGTGGAGCCAGAGCAAGAGAGAGAGAG 823
Db      541 CTGGGGGTGAAAGCTCAACCGTGGGCTCTGAGCGGAGCCAGAGTAAAGAGAGAGAGAG 600
QY      824 AATCAGGCTGACCTGAGAGCGGAGCCCGGAGAGAGAGCAAGCCCACTGAGAGAGCTG 883
Db      601 AATCAGGCTGACCTGAGAGCGGAGCCCGGAGAGAGAGCAAGCCCGTGGAGAGCTG 660
QY      884 CGGAGCTGGAGAGAGAGAGAGAGAGCCCGCTCAGGTTCCAGGAGTGAACGCGCTGAG 943
Db      661 CGTGGCTGGAGAGAGAGAGAGAGAGCCCACTCAGGTTCCAGGAGTGAACGCGCTGAG 720
QY      944 CAGTCTGGCTGCTACCACTTGTGCTGATGAGAGACATCGAGAGAGAGAGAGAGAGAG 1003
Db      721 CAGCAGAGCTGCTACCACTTGTGCTGATGAGAGACATGAGAGAGAGAGAGAGAGAGAG 780

```


QY 1004 GATCTCCGGGATAGAAATACACGCTCCCAATTTGGGGCTGGCTCCCTTCCTGGGACC 1063
DB 781 GACCTGGCGGGATAGAAATACACGCTCCCAATTTGGGGCTGGCTCCCTTCCTGGGACC 840
QY 1064 CAGAGTTCAGAACTGCCCCCGGACCTCAATCCCTGATCTCGCTCCCATGAGCCG 1123
DB 841 CAGAGTTCAGAGCTGCCCCCGGACCTCAATCCCTGATCTCCCAATCCCTGATCTCCCATGAGCCG 900
QY 1124 GAAGCCATCCCATCCCAACCGAAAGCCCAAGGCGGTGACCCGGCTCTCTTCCACTTC 1183
DB 901 GAAGCTCCCATCCCAACCGAAAGCCCAAGGCGGTGACCCGGCTCTCTTCCACTTC 960
QY 1184 CTTCAGACCCCAATCGCAAGGTCTCAGAGCTCCAGCAACGGCTCCGGGGACACCAAGAC 1243
DB 961 CTTCAGACCCCAATTCGCAAGGTCTCAGAGCTCCAGCAACGGCTCCGGGGACACCAAGAT 1020
QY 1244 GGGAGCAAGCACTTTGTAGAGTCCCAAGGCGGCAAGGTGTGGGTGTGGGCGAC 1303
DB 1021 GGGAGCAAGCACTTTGTAGAGTCCCAAGGCGGCAAGATGGGTATGGGCGAC 1080
QY 1304 GTGGCCAGAGGGGCAAGAAACAGCCCTCTGGGACCCCGCATTCCTCGGTGTCCGCC 1363
DB 1081 GGGGCGAGAGGTGCAAGAGCAAGCTCCACTGTGACCAACCACTACTGTCTCCGCC 1140
QY 1364 TCCGCCACCTGCTGCGAGCGCGGCTCTCTCCCTCTGAGCCCGCTGGGCAAG 1423
DB 1141 TCTGCCATCTGGGCGAGCGGCTCTCTCCCACTGGGACCCCTGGGCGACAG 1200
QY 1424 AAGCACAAGCAACCGAGCCAGAGAGCCAGAGGCTGCGGGGCTGCAAGCAACTG 1483
DB 1201 AAACACAAGCATCGAGCCAGAGAGCCAGGCGAGCTGCGGGGCTGCAAGGCGCCCTG 1260
QY 1484 GCCTCAGGTGG---CCCTGTCTTGGGGCGGAGACCTGCGGGAGCTGCCGCTTGGTG 1540
DB 1261 GCTGCAAGAGGCTCCACGCTCATGAGGCGGAGAGGCTGAGGGAGCTGCGGTGTG 1320
QY 1541 GTGTATGAGAGCGAGCGCGGAGCGGATCCAGAGCATGAGCAGCAACCACTATGAA 1600
DB 1321 GTGTACAGAGCCAGGCTGGGCGAGCGGCTCCAGAGCAAGCAACCACTATGACCAACCA 1380
QY 1601 CATCACCACCATTTACCAACAC 1621
DB 1381 CATCACCACCATTTACCAAC 1401

RESULT 6
AAL03214
ID AAL03214 standard; DNA; 2379 BP.
XX AAL03214;
AC AAL03214;
DT 21-NOV-2001 (first entry)
XX Human reproductive system related antigen DNA SEQ ID NO: 5902.
DE Human reproductive system related antigen; reproductive system disorder;
KM cancer; gene therapy; ds.
XX Homo sapiens.
OS Homo sapiens.
XX W0200155320-A2.
PN W0200155320-A2.
PD 02-AUG-2001.
XX 02-AUG-2001.
PF 17-JAN-2001; 2001WO-US001339.
XX 17-JAN-2001; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.

20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249254P.
PR 17-NOV-2000; 2000US-0249255P.
PR 17-NOV-2000; 2000US-0249256P.
PR 17-NOV-2000; 2000US-0249257P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
(HUMA-) HUMAN GENOME SCT INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-465570/50.
Isolated nucleic acid molecule encoding a reproductive system antigen is
used in preventing, treating or ameliorating a medical condition.
Disclosure; SEQ ID NO 5902; 1297bp + Sequence Listing; English.
The present invention provides the protein and coding sequences of a
number of human reproductive system related antigens. These can be used
in the prevention and treatment of reproductive system disorders,
including cancer. The present sequence is a genomic sequence encoding a
protein of the invention

XX Sequence 2379 BP; 548 A; 689 C; 539 G; 603 T; 0 U; 0 Other;
SQ Query Match 42.1%; Score 783; DB 4; Length 2379;
Best Local Similarity 100.0%; Pred.No.1e-146;
Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1046 GGCTCCCTCCGTTGCGGCGCCGAGAGTCAGAACTGCCCCCGGCACTTCATCCACTCGA 1105
DB 840 GGCTCCCTCCCTTCGTTGCGGCGCCGAGAGTCAGAACTGCCCCCGGCACTTCATCCACTCGA 899
QY 1106 TCTGCTCCCATGAGCCGGAAGCATTCACATCCACAGAAAGCCCAAGGCTGGAC 1165
DB 900 TCTGCTCCCATGAGCCGGAAGCATTCACATCCACAGAAAGCCCAAGGCTGGAC 959
QY 1166 CCGGCTCTCTTCACATCTCTTGAACCCCAATGCCAAGGCTCAGAGCTCCAGCAAG 1225
DB 960 CCGGCTCTCTTCACATCTCTTGAACCCCAATGCCAAGGCTCAGAGCTCCAGCAAG 1019
QY 1226 CTCGGGGGACCCAGAGCGGGAGCAAGCACTTGTAGGTCCTCCCAAGGCCAGGCGAAG 1285
DB 1020 CTCGGGGGACCCAGAGCGGGAGCAAGCACTTGTAGGTCCTCCCAAGGCCAGGCGAAG 1079
QY 1286 AGTGTGGGTGTGGGCAAGTGGCCAGAGGGGCAAGAAACAAGCTCCCTTGGGACCGGC 1345
DB 1080 AGTGTGGGTGTGGGCAAGTGGCCAGAGGGGCAAGAAACAAGCTCCCTTGGGACCGGC 1139
QY 1346 ATCCCTGCGGTGTCCTCCCTCCGCACTGGCTGCGAGCCCGGCTCTCCCTCCCTA 1405
DB 1140 ATCCCTGCGGTGTCCTCCCTCCGCACTGGCTGCGAGCCCGGCTCTCCCTCCCTA 1199
QY 1406 GCCCGCCCTCGGGGCAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1465
DB 1200 GCCCGCCCTCGGGGCAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1259
QY 1466 GGCCTGAGGACCACTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 1525
DB 1260 GGCCTGAGGACCACTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 1319
QY 1526 CTGCGCGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1585
DB 1320 CTGCGCGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1379
QY 1586 CACCAACCATGAACATCAACCACTTACCACTTACCACTTACCACTTACCACTTACCA 1645
DB 1380 CACCAACCATGAACATCAACCACTTACCACTTACCACTTACCACTTACCACTTACCA 1439
QY 1646 CCAAGGCGCCACCTGCAATGAAAGACCCACCCCGACACCAAGGCAATTATT 1705
DB 1440 CCAAGGCGCCACCTGCAATGAAAGACCCACCCCGACACCAAGGCAATTATTATT 1499
QY 1706 CTATTATTATTGTTATATGATGATGATGATGATGATGATGATGATGATGATGAT 1765
DB 1500 CTATTATTATTGTTATATGATGATGATGATGATGATGATGATGATGATGATGAT 1559
QY 1766 ATTTAGCTACCTACATGATGATGATGATGATGATGATGATGATGATGATGAT 1825
DB 1560 ATTTAGCTACCTACATGATGATGATGATGATGATGATGATGATGATGATGAT 1619
QY 1826 GTT 1828
DB 1620 GTT 1622
RESULT 7
AAL03216
ID AAL03216 standard; DNA; 2379 BP.
XX
AC AAL03216;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 5904.

XX Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX MO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001, 2001MO-US001339.
XX
PR 31-JAN-2000, 2000US-0179065P.
PR 04-FEB-2000, 2000US-0180628P.
PR 24-FEB-2000, 2000US-0184664P.
PR 02-MAR-2000, 2000US-0186350P.
PR 16-MAR-2000, 2000US-0189874P.
PR 17-MAR-2000, 2000US-0190076P.
PR 18-APR-2000, 2000US-0198123P.
PR 19-MAY-2000, 2000US-0205515P.
PR 07-JUN-2000, 2000US-0209467P.
PR 28-JUN-2000, 2000US-0214886P.
PR 30-JUN-2000, 2000US-0215135P.
PR 07-JUL-2000, 2000US-0216647P.
PR 11-JUL-2000, 2000US-0216880P.
PR 11-JUL-2000, 2000US-0217487P.
PR 11-JUL-2000, 2000US-0217496P.
PR 14-JUL-2000, 2000US-0218290P.
PR 26-JUL-2000, 2000US-0220963P.
PR 26-JUL-2000, 2000US-0220964P.
PR 14-AUG-2000, 2000US-0224518P.
PR 14-AUG-2000, 2000US-0224519P.
PR 14-AUG-2000, 2000US-0225213P.
PR 14-AUG-2000, 2000US-0225214P.
PR 14-AUG-2000, 2000US-0225266P.
PR 14-AUG-2000, 2000US-0225267P.
PR 14-AUG-2000, 2000US-0225268P.
PR 14-AUG-2000, 2000US-0225270P.
PR 14-AUG-2000, 2000US-0225447P.
PR 14-AUG-2000, 2000US-0225757P.
PR 14-AUG-2000, 2000US-0225758P.
PR 14-AUG-2000, 2000US-0225759P.
PR 18-AUG-2000, 2000US-0226279P.
PR 22-AUG-2000, 2000US-0226681P.
PR 22-AUG-2000, 2000US-0226686P.
PR 22-AUG-2000, 2000US-0227182P.
PR 23-AUG-2000, 2000US-0227009P.
PR 30-AUG-2000, 2000US-0228984P.
PR 01-SEP-2000, 2000US-0229287P.
PR 01-SEP-2000, 2000US-0229343P.
PR 01-SEP-2000, 2000US-0229344P.
PR 01-SEP-2000, 2000US-0229345P.
PR 05-SEP-2000, 2000US-0229509P.
PR 05-SEP-2000, 2000US-0229513P.
PR 06-SEP-2000, 2000US-0230437P.
PR 06-SEP-2000, 2000US-0230438P.
PR 08-SEP-2000, 2000US-0231242P.
PR 08-SEP-2000, 2000US-0231243P.
PR 08-SEP-2000, 2000US-0231244P.
PR 08-SEP-2000, 2000US-0231413P.
PR 08-SEP-2000, 2000US-0231414P.
PR 08-SEP-2000, 2000US-0232080P.
PR 08-SEP-2000, 2000US-0232081P.
PR 12-SEP-2000, 2000US-0231968P.
PR 14-SEP-2000, 2000US-0232397P.
PR 14-SEP-2000, 2000US-0232398P.
PR 14-SEP-2000, 2000US-0232399P.
PR 14-SEP-2000, 2000US-0232400P.
PR 14-SEP-2000, 2000US-0232401P.
PR 14-SEP-2000, 2000US-0233063P.
PR 14-SEP-2000, 2000US-0233064P.
PR 14-SEP-2000, 2000US-0233065P.
PR 21-SEP-2000, 2000US-0234223P.

PR 21-SEP-2000, 2000US-0234374P.
PR 25-SEP-2000, 2000US-0234997P.
PR 25-SEP-2000, 2000US-0234998P.
PR 26-SEP-2000, 2000US-0235484P.
PR 27-SEP-2000, 2000US-0235834P.
PR 27-SEP-2000, 2000US-0235836P.
PR 29-SEP-2000, 2000US-0236327P.
PR 29-SEP-2000, 2000US-0236367P.
PR 29-SEP-2000, 2000US-0236368P.
PR 29-SEP-2000, 2000US-0236369P.
PR 29-SEP-2000, 2000US-0236370P.
PR 02-OCT-2000, 2000US-0236802P.
PR 02-OCT-2000, 2000US-0237037P.
PR 02-OCT-2000, 2000US-0237038P.
PR 02-OCT-2000, 2000US-0237039P.
PR 02-OCT-2000, 2000US-0237040P.
PR 13-OCT-2000, 2000US-0239935P.
PR 13-OCT-2000, 2000US-0239937P.
PR 20-OCT-2000, 2000US-0240960P.
PR 20-OCT-2000, 2000US-0241221P.
PR 20-OCT-2000, 2000US-0241785P.
PR 20-OCT-2000, 2000US-0241786P.
PR 20-OCT-2000, 2000US-0241787P.
PR 20-OCT-2000, 2000US-0241808P.
PR 20-OCT-2000, 2000US-0241809P.
PR 20-OCT-2000, 2000US-0241826P.
PR 01-NOV-2000, 2000US-0244617P.
PR 08-NOV-2000, 2000US-0246474P.
PR 08-NOV-2000, 2000US-0246475P.
PR 08-NOV-2000, 2000US-0246476P.
PR 08-NOV-2000, 2000US-0246477P.
PR 08-NOV-2000, 2000US-0246478P.
PR 08-NOV-2000, 2000US-0246523P.
PR 08-NOV-2000, 2000US-0246524P.
PR 08-NOV-2000, 2000US-0246525P.
PR 08-NOV-2000, 2000US-0246526P.
PR 08-NOV-2000, 2000US-0246527P.
PR 08-NOV-2000, 2000US-0246528P.
PR 08-NOV-2000, 2000US-0246532P.
PR 08-NOV-2000, 2000US-0246609P.
PR 08-NOV-2000, 2000US-0246610P.
PR 08-NOV-2000, 2000US-0246611P.
PR 08-NOV-2000, 2000US-0246613P.
PR 17-NOV-2000, 2000US-0249207P.
PR 17-NOV-2000, 2000US-0249208P.
PR 17-NOV-2000, 2000US-0249209P.
PR 17-NOV-2000, 2000US-0249210P.
PR 17-NOV-2000, 2000US-0249211P.
PR 17-NOV-2000, 2000US-0249212P.
PR 17-NOV-2000, 2000US-0249213P.
PR 17-NOV-2000, 2000US-0249214P.
PR 17-NOV-2000, 2000US-0249215P.
PR 17-NOV-2000, 2000US-0249216P.
PR 17-NOV-2000, 2000US-0249217P.
PR 17-NOV-2000, 2000US-0249218P.
PR 17-NOV-2000, 2000US-0249244P.
PR 17-NOV-2000, 2000US-0249245P.
PR 01-DEC-2000, 2000US-0250391P.
PR 05-DEC-2000, 2000US-0251030P.
PR 05-DEC-2000, 2000US-0251988P.
PR 05-DEC-2000, 2000US-0256719P.
PR 06-DEC-2000, 2000US-0251479P.
PR 06-DEC-2000, 2000US-0251856P.
PR 06-DEC-2000, 2000US-0251868P.
PR 06-DEC-2000, 2000US-0251869P.
PR 06-DEC-2000, 2000US-0251989P.
PR 08-DEC-2000, 2000US-0251990P.


```

QY 1227 TCCGGGGCAACCGAGACGGGAGCACTTGTGAGGTCCCGGAGGCCCAAGGCGAAGA 1286
DB 181 TCCGGGGCAACCGAGACGGGAGCACTTGTGAGGTCCCGGAGGCCCAAGGCGAAGA 240
QY 1287 GTGTGGGTGTGGGCGCAGTGGGCGAGAGGCGAAGAAACAAGCCCTCTGTGGACCCGCGCA 1346
DB 241 GTGTGGGTGTGGGCGCAGTGGGCGAGAGGCGAAGAAACAAGCCCTCTGTGGACCCGCGCA 300
QY 1347 TCCCTGCGGTGTCCCTCTCGGCCACCTGCGTGGCAGCCCGGCTCTCTCCCTCTAG 1406
DB 301 TCCCTGCGGTGTCCCTCTCGGCCACCTGCGTGGCAGCCCGGCTCTCTCCCTCTAG 360
QY 1407 CCCCCCTTGGGGCAAGAGAACCAAGACCGGAGCGAAGAGGCGAAGGCGTGGCGGG 1466
DB 361 CCCCCCTTGGGGCAAGAGAACCAAGACCGGAGCGAAGGCGAAGGCGTGGCGGG 420
QY 1467 GCCTGAGGACCACTGAGCTCAGGTGGCCCTGTCTGGGGGCGGAGACCTGCGGAGG 1526
DB 421 GCCTGAGGACCACTGAGCTCAGGTGGCCCTGTCTGGGGGCGGAGACCTGCGGAGG 480
QY 1527 TGGCCGCTTGTGTGTATGAGAGCCAGGCCGCGGCAAGCGTTCAGAGACATGAGGACC 1586
DB 481 TGGCCGCTTGTGTGTATGAGAGCCAGGCCGCGGCAAGCGTTCAGAGACATGAGGACC 540
QY 1587 ACCACCACTGAAACATCACCACCATTAACCACTTCTACAGACATAG 1636
DB 541 ACCACCACTGAAACATCACCACCATTAACCACTTCTACAGACATAG 590

RESULT 9
AAS92082
ID AAS92082 standard; cDNA, 1561 BP.
XX
AC AAS92082;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #27886.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PE 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT,
XX
PI WPI; 2001-639362/73.
XX
PI P-PSDB; ABG27895.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 27886; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal

```

```

CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1561 BP; 396 A; 459 C; 397 G; 309 T; 0 U; 0 Other;
XX
Query Match 24.5%; Score 455.8; DB 5; Length 1561;
Best Local Similarity 99.6%; Pred. No. 2,5e-81;
Matches 457; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 591 AGCTCCAGTSCAGTGTCTCCATGAGAGGACAGCCGCGCAGAGTGAACCTTCACTGT 650
DB 1 AGCTCCAGTSCAGTGTCTCCATGAGAGGACAGCCGCGCAGAGTGAACCTTCACTGT 60
QY 651 ATGACTTTGACAAACAAGGAGGTCAACCGAGAGACATACCACTTCTGACACCA 710
DB 61 ATGACTTTGACAAACAAGGAGGTCAACCGAGAGACATACCACTTCTGACACCA 120
QY 711 TCTATGAGTGTGTGACTCTCTGTCAACCACTTCTGACCAATCCAGAAATGTGGGG 770
DB 121 TCTATGAGTGTGTGACTCTCTGTCAACCACTTCTGACCAATCCAGAAATGTGGGG 180
QY 771 TAAAGCTCACCGTGGCCCGCCGATGGAGCGCAGAGCAAGAGAGTCTTGTCAATCAG 830
DB 181 TAAAGCTCACCGTGGCCCGCCGATGGAGCGCAGAGCAAGAGAGTCTTGTCAATCAG 240
QY 831 CTGACCTGACAGACCGCAAGGCCCGCAGACAGACCAAGCCACTGAGACCTTGGAGCT 890
DB 241 CTGACCTGACAGACCGCAAGGCCCGCAGACAGACCAAGCCACTGAGACCTTGGAGCT 300
QY 891 GGGAGAAAGAGACGACGAGCCCGCTCAAGTTCAGAGGTGACAGCCGCTGAGACATCTG 950
DB 301 GGGAGAAAGAGACGACGAGCCCGCTCAAGTTCAGAGGTGACAGCCGCTGAGACATCTG 360
QY 951 GCTGCTACGACCAATGGGTATGATGAGACATGAGAGAGAAACCACTAATTAGATCTG 1010
DB 361 GCTGCTACGACCAATGGGTATGATGAGACATGAGAGAGAAACCACTAATTAGATCTG 420
QY 1011 CCGGAGATAGAAACTACACGTTCCCAATTTGGGCTGCT 1049
DB 421 CCGGAGATAGAAACTACACGTTCCCAATTTGGGCTGCT 459

RESULT 10
ACH21474
ID ACH21474 standard; cDNA, 427 BP.
XX
AC ACH21474;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human adult liver cDNA #1086.
XX
KW Human; ss; sequencing by hybridisation; SSH; expressed sequence tag; EST;
XX
KW genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.
XX
PN US2003073623-A1.
XX
PD 17-APR-2003.

```

PF		30-Jul-2001; 2001US-00918995.
XX		
PR		30-Jul-2001; 2001US-00918995.
XX		
PA	(DMAA/) DRMANAC R T.	
PA	(LABAT I.) LABAT I.	
PA	(STAC/) STACHE-CRAIN B.	
PA	(DICK/) DICKSON M C.	
PA	(JONE/) JONES L W.	
XX		
PI	Dmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;	
DR		
XX	WPI; 2003-615964/58.	
PT	New polynucleotide sequences obtained from various cDNA libraries, useful	
PT	as hybridization probes, as oligomers for PCR, for chromosome and gene	
PT	mapping, in the recombinant production of protein, or in generating	
PT	antisense DNA or RNA.	
XX		
PS	Claim 1; SEQ ID NO 8686; 44pp; English.	
CC	The invention relates to an isolated polynucleotide comprising any one of	
CC	38043 CDNA sequences, appearing as ACH12789-AHG50831, whose sequence was	
CC	determined by the technique of SBH (sequencing by hybridisation). Also	
CC	included is a purified polypeptide comprising a sequence corresponding to	
CC	a reading frame of the novel polynucleotide. The nucleic acid sequences	
CC	are useful in diagnostics as expressed sequence tags (EST) for	
CC	identifying expressed genes or for physical mapping of the human genome,	
CC	in forensics, in assessing biodiversity, or in identifying mutations	
CC	responsible for genetic disorders and other traits. The nucleotide	
CC	sequences are also useful as hybridisation probes, as oligomers for PCR,	
CC	for chromosome and gene mapping, in the recombinant production of	
CC	protein, or in generating antisense DNA or RNA. The purified polypeptide	
CC	is useful for generating antibodies specific for it. The present sequence	
CC	is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data	
CC	for this patent did not form part of the printed specification, but was	
CC	obtained in electronic format directly from USFTO at	
CC	seqdata.uspto.gov/sequence.html?DocID=20030073623	
XX		
SQ	Sequence 427 BP; 84 A; 177 C; 112 G; 53 T; 0 U; 1 Other;	
	Query Match 22.5%; Score 418.2; DB 9; Length 427;	
	Best Local Similarity 99.3%; Pred. No. 5.6e-74;	
	Matches 420; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
OY	1076 CTGCCCCCGCAGCTCCCATTCCTCCAGTCTGCCTCCCAATGAGCCGGAAAGCATCCAC	1135
DB	4 CTGCCCCCGCAGCTCCCATTCCTCCAGTCTGCCTCCCAATGAGCCGGAAAGCATCCAC	63
OY	1136 ATCCACACCGAAAGCCCAAAGGCGGTGAACC CGGCGCTCTTCACTTGACACCCCA	1195
DB	64 ATCCACACCGAAAGCCCAAAGGCGGTGAACC CGGCGCTCTTCACTTGACACCCCA	123
OY	1196 ATGCGCAAGTCTCAGAGCTCCAGCAACGCGCTCCGGGACACCAAGACGGACAAGAC	1255
DB	124 ATGCGCAAGTCTCAGAGCTCCAGCAACGCGCTCCGGGACACCAAGACGGACAAGAC	183
OY	1256 TTGTGAGAGTCCCCCAAGGCCCAAGGCAAGAATGTGGTGTGGCAAGTGGCCCAAGAGG	1315
DB	184 TTGTGAGAGTCCCCCAAGGCCCAAGGCAAGAATGTGGTGTGGCAAGTGGCCCAAGAGG	243
OY	1316 GCAGAAACAAGCCCCCTCTGGGAAACCGGCATACCTGTGGGATGCCCTCCGCGCCACTG	1375
DB	244 GCAGAAACAAGCCCCCTCTGGGAAACCGGCATACCTGTGGGATGCCCTCCGCGCCACTG	303
OY	1376 GCTGCGAACCCGGCCCTCTCCCTCCTTAGCCCCCCTTGCGGACACAAGACACAAGAC	1435
DB	304 GCTGCGAACCCGGCCCTCTCCCTCCTTAGCCCCCCTTGCGGACACAAGACACAAGAC	363
OY	1436 CGAGTCGAAGAGAGCCAGACGGGCTGGCCGGGCGCTGACAGCACCATGGCTCAGTGGC	1495
DB	364 CGAGTCGAAGAGAGCCAGGCTGGCCGGGCGCTGACAGCACCATGGCTCAGTGGC	423

QY 1496 CCT 1498
|||
Db 424 CCT 426

RESULT 11
AA\$78752 ID AAS78752 standard; cDNA, 1743 BP.
XX AC AAS78752;
XX DT 13-FEB-2002 (first entry)
XX DE DNA encoding novel human diagnostic protein #14556.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI: 2001-639362/73.
XX DR P-PBDB; ABG14565.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

PS Claim 1; SEQ ID NO 14556; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridization probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AA\$94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pcc_sequences
XX SQ Sequence 1743 BP; 367 A; 562 C; 511 G; 303 T; 0 U; 0 Other;

Query Match 15.3%; Score 284.2; DB 5; Length 1743;
Best Local Similarity 77.8%; Pred. No. 5.4e-47;
Matches 410; Conservative 0; Mismatches 3; Indels 114; Gaps 1;

1047 GCTCCCTTCCTCGTGGCCAGAGTCAGAACTGCCCCCCCGCAGCACTCATATCCACATCGAT 1106

```
Db 332 GCTCCCTTCCTGCGGCGCCAGAGTCAAGATCTGCCCCCCCCCGCACTTCAATCCACTTGAT 391
Qy 1107 CTGCTCCCATGAGCGCGGAGCCATCCATCCCAACCGAAGGCCCAAGCGCTGAGCC 1166
Db 392 CTGCTCCCATGAGCGCGGAGCCATCCATCCCAACCGAAGGCCCAAGCGCTGAGCC 451
Qy 1167 CGGCTCTCTTCCACTTCTCTTGAACCCCAATCCGCAAGGTCTCAGAGCTCCAGCAACGGC 1226
Db 452 CGGCTCTCTTCCACTTCTCTTGAACCCCAATCCGCAAGGTCTCAGAGCTCCAGCAACGGC 511
Qy 1227 TCCGGGGCAACCCGAGACGGAGAGCACTTTGAGGTCCCGCCCAAGGCCCAAGGCAAGA 1286
Db 512 TCCGGGGCAACCCGAGACGGAGAGCACTTTGAGGTCCCGCCCAAGGCCCAAGGCAAAA 571
Qy 1287 GTGTGGGTGTGGGCAAGTGGCCAGAGGGGCAAGAAACAAAGCCCTCTGGAACCCGCCA 1346
Db 572 GTGTG-----G 576
Qy 1347 TCCCTGCGGTGTCCCTCCCTCCGCCCACTGAGCTGCGAGCCCGCCCTCTCCCTCTAG 1406
Db 577 -----G 577
Qy 1407 CCCCCCTCGGGCAAGAGACACAGACCGAGCCAGAGAGCGAGCGGCTGCCGG 1466
Db 578 CCCCCCTCGGGCAAGAGACACAGACCGAGCCAGAGAGCGAGCGGCTGCCGG 637
Qy 1467 GCCTGACGAGCAACCACTGAGCTTCAAGTGGCCCTGCTCTGAGGCGGAGCACTTCGGAGC 1526
Db 638 GCCTGACGAGCAACCACTGAGCTTCAAGTGGCCCTGCTCTGAGGCGGAGCACTTCGGAGC 697
Qy 1527 TGGCCGCTTGTGTGTATGAGAGCGAGCCGAGCGAGCCGCTTCAG 1573
Db 698 TGGCCGCTTGTGTGTATGAGAGCGAGCCGAGCGAGCCGCTTCAG 744
```

RESULT 12

AAS92081 ID AAS92081 standard; cDNA; 1743 BP.

AAS92081; AC XX

13-FEB-2002 (first entry) DT XX

DNA encoding novel human diagnostic protein #27885. DE XX

Human; chromosome mapping; gene mapping; gene therapy; forensic; KM XX

food supplement; medical imaging; diagnostic; genetic disorder; ss. XX

Homo sapiens. OS XX

W0200175067-A2. PN XX

11-OCT-2001. PD XX

30-MAR-2001; 2001MO-US008631. PF XX

31-MAR-2000; 2000US-00540217. PR XX

23-AUG-2000; 2000US-00649167. PX XX

(HYSE-) HYSEQ INC. PA XX

Drmanac RT, Liu C, Tang YT; PI XX

WPI; 2001-639362/73. DR XX

P-PSDB; ABG27894. DR XX

New isolated polynucleotide and encoded polypeptides, useful in PT XX

diagnostics, forensics, gene mapping, identification of mutations PT XX

responsible for genetic disorders or other traits and to assess PT XX

PS

```
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probe, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences
CC XX
```

SQ Sequence 1743 BP; 367 A; 562 C; 511 G; 303 T; 0 U; 0 Other;

Query Match 15.3%; Score 284.2; DB 5; Length 1743;

Best Local Similarity 77.8%; Pred. No. 5,4e-47;

Matches 410; Conservative 0; Mismatches 3; Indels 114; Gaps 1;

```
Qy 1047 GCTCCCTTCCTGCGGCGCCAGAGTCAAGATCTGCCCCCCCCCGCACTTCAATCCACTTGAT 1106
Db 332 GCTCCCTTCCTGCGGCGCCAGAGTCAAGATCTGCCCCCCCCCGCACTTCAATCCACTTGAT 391
Qy 1107 CTGCTCCCATGAGCGCGGAGCCATCCATCCCAACCGAAGGCCCAAGCGCTGAGCC 1166
Db 392 CTGCTCCCATGAGCGCGGAGCCATCCATCCCAACCGAAGGCCCAAGCGCTGAGCC 451
Qy 1167 CGGCTCTCTTCCACTTCTCTTGAACCCCAATCCGCAAGGTCTCAGAGCTCCAGCAACGGC 1226
Db 452 CGGCTCTCTTCCACTTCTCTTGAACCCCAATCCGCAAGGTCTCAGAGCTCCAGCAACGGC 511
Qy 1227 TCCGGGGCAACCCGAGACGGAGAGCACTTTGAGGTCCCGCCCAAGGCCCAAGGCAAGA 1286
Db 512 TCCGGGGCAACCCGAGACGGAGAGCACTTTGAGGTCCCGCCCAAGGCCCAAGGCAAAA 571
Qy 1287 GTGTGGGTGTGGGCAAGTGGCCAGAGGGGCAAGAAACAAAGCCCTCTGGAACCCGCCA 1346
Db 572 GTGTG-----G 576
Qy 1347 TCCCTGCGGTGTCCCTCCCTCCGCCCACTGAGCTGCGAGCCCGCCCTCTCCCTTAG 1406
Db 577 -----G 577
Qy 1407 CCCCCCTCGGGCAAGAGACACAGACCGAGCCAGAGAGCGGCTGCCGG 1466
Db 578 CCCCCCTCGGGCAAGAGACACAGACCGAGCCAGAGAGCGGCTGCCGG 637
Qy 1467 GCCTGACGAGCAACCACTGAGCTTCAAGTGGCCCTGCTCTGAGGCGGAGCACTTCGGAGC 1526
Db 638 GCCTGACGAGCAACCACTGAGCTTCAAGTGGCCCTGCTCTGAGGCGGAGCACTTCGGAGC 697
Qy 1527 TGGCCGCTTGTGTGTATGAGAGCGAGCCGAGCGAGCCGCTTCAG 1573
Db 698 TGGCCGCTTGTGTGTATGAGAGCGAGCCGAGCGAGCCGCTTCAG 744
```

RESULT 13

ADC31966 ID ADC31966 standard; cDNA; 1743 BP.

ADC31966; AC XX

18-DEC-2003 (first entry) DT XX

XX Human novel cDNA contig sequence, SEQ ID NO:2048.
DE
XX Human, diagnostic, drug screening, forensics, gene mapping;
XX biodefensibility assessment; Parkinson's disease; Alzheimer's disease;
XX neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
XX ulcer; osteoporosis; autoimmune disease; cancer;
XX molecular weight marker; food supplement; antiparkinsonian; nootropic;
XX neuroprotective; antianemic; anticoagulant; thrombolytic; vulnerary;
XX anticancer; osteoprotic; immunosuppressive; antiinflammatory; cyostatic;
XX gene therapy; ss.
XX Homo sapiens.
OS
XX W02003029271-A2.
PN
XX 10-APR-2003.
PD
XX 24-SEP-2002; 2002WO-US030474.
PE
XX 24-SEP-2001; 2001US-0324631P.
PR
XX (HYSE-) HYSEQ INC.
PA
XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
PI Haley-Vicente D, Drmanac RT;
XX WPI; 2003-371981/35.
DR P-PSDB; ADC32733.
XX
PT New polynucleotide and polypeptide useful for diagnosing, preventing or
PT treating conditions such as neurodegenerative diseases, anemias, platelet
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
PT cancer.
XX
XX Example 2; SEQ ID NO 2048; 1185bp; English.
XX
XX The invention relates to 971 novel human cDNA sequences (ADC29919-
CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
CC invention also relates to nucleic acid sequences over 99% identical with
CC the novel human cDNAs. The invention additionally encompasses expression
CC vectors and host cells comprising a nucleic acid of the invention; the
CC recombinant production of a polypeptide of the invention; an antibody
CC against a polypeptide of the invention; a method of detecting
CC polynucleotides or polypeptides of the invention; and methods of
CC identifying a compound which binds to a polypeptide of the invention. The
CC invention further discloses methods of preventing, treating or
CC ameliorating a medical condition; kits comprising polynucleotide probes
CC and/or monoclonal antibodies for carrying out the methods of the
CC invention; methods for the identification of compounds that modulate the
CC expression or activity of the polynucleotide and/or polypeptide; and 767
CC contig sequences corresponding to the cDNA sequences of the invention
CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
CC -ADC33394). The nucleic acids and polypeptides of the invention are
CC useful in diagnostics, drug screening, forensics, gene mapping, in the
CC identification of mutations responsible for genetic disorders or other
CC traits, for assessing biodiversity, and in producing many other types of
CC data and products dependent on DNA and amino acid sequences. They are
CC also used for treating diseases such as Parkinson's disease, Alzheimer's
CC disease and other neurodegenerative diseases, anaemia, platelet
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primers, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a human contig
CC sequence used in an example of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 1743 BP; 367 A; 562 C; 511 G; 303 T; 0 U; 0 Other;

Query Match 15.3%; Score 284.2; DB 10; Length 1743;
Best Local Similarity 77.8%; Pred. No. 5,4e-47;
Matches 410; Conservative 0; Mismatches 3; Indels 114; Gaps 1;
QY 1047 GCTCCCTTCCTGCGGAGCCAGAGTCAAGTCTGCCCCCGACCTCCATCCACTGAT 1106
DB 332 GCTCCCTTCCTGCGGAGCCAGAGTCAAGTCTGCCCCCGACCTCCATCCACTGAT 391
QY 1107 CTGCTCCCTGAGCCGAGGAGCCATCCATCCCAAGCCGAGGAGGCTGAGC 1166
DB 392 CTGCTCCCTGAGCCGAGGAGCCATCCATCCCAAGCCGAGGAGGCTGAGC 451
QY 1167 CGGCTTCCTTCACCTTCCTTGACACCCCAATCCGAGTCTGAGCTCCAGCAAGC 1226
DB 452 CGGCTTCCTTCACCTTCCTTGACACCCCAATCCGAGTCTGAGCTCCAGCAAGC 511
QY 1227 TCCGAGGACCCAGAGCCGAGGAGCACTTTGTAGTCTCCCAAGGAGCCAGGCAAGA 1286
DB 512 TCCGAGGACCCAGAGCCGAGGAGCACTTTGTAGTCTCCCAAGGAGCCAGGCAAGA 571
QY 1287 GTGTGGGTGTGGGCACGTGGCCAGAGGGGCAAGAAACAGCCCTCTGGAGCCGCCA 1346
DB 572 GTGTG----- 576
QY 1347 TCCCTGCGGTGTCCTCCCTCCGCCCACTGAGTCCAGCCGAGCTCTCCCTCCTAG 1406
DB 577 -----G 577
QY 1407 CCCCCTCGGACCAAGAGCAGACAGCAGCAGCAAGAGAGAGCCAGAGGCTGCCGAG 1466
DB 578 CCCCCTCGGACCAAGAGCAGACAGCAGCAGCAGAGAGAGCCAGAGGCTGCCGAG 637
QY 1467 GCCTGACAGCAGCAGCAGTCCCTGAGTGGCCCTGCTGCTGAGGAGGAGCAGCAGTGGGAGC 1526
DB 638 GCCTGACAGCAGCAGCAGTCCCTGAGTGGCCCTGCTGCTGAGGAGGAGCAGCAGTGGGAGC 697
QY 1527 TGCCCGCTGTGTGTATGAGAGCCAGGCGGAGCCGATCCAG 1573
DB 698 TGCCCGCTGTGTGTATGAGAGCCAGGCGGAGCCGATCCAG 744
RESULT 14
AAS27117
ID AAS27117 standard; cDNA; 1966 BP.
AC AAS27117;
XX
DT 07-NOV-2001 (first entry)
XX
XX cDNA encoding novel signal transduction pathway protein, Seq ID 152.
DE
XX Neuroprotective; cyostatic; dermatological; immunosuppressive; tumour;
XX antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
XX immune system disorder; rheumatoid arthritis; inflammatory condition;
XX organ transplant rejection; infection; hepatitis C; blood disorder;
XX sickle cell anemia; hyperproliferative disorder; Gaucher's disease;
XX neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
XX chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
XX cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
XX reproductive system; gastrointestinal; liver disorder; AIDS; ss;
XX acquired immune deficiency syndrome.
XX
XX Homo sapiens.
OS
XX W0200154733-A1.
PN
XX 02-AUG-2001.
PD
XX 17-JAN-2001; 2001WO-US001312.
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229345P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232377P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235464P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236377P.
PR 29-SEP-2000; 2000US-0236377P.
PR 29-SEP-2000; 2000US-0236386P.
PR 29-SEP-2000; 2000US-0236389P.
PR 29-SEP-2000; 2000US-0236397P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.

PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 06-DEC-2000; 2000US-0251656P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251899P.
PR 08-DEC-2000; 2000US-0251907P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
XX
PI
XX WPI: 2001-465460/50.
XX P-PDB; AAU17200.
DR
DR
XX
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders.
XX
PS Claim 1; SEQ ID NO 152; 880bp; English.

XX The invention relates to novel isolated polypeptides (I), and
 CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
 CC diagnosing, preventing and treating diseases including immune system
 CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
 CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
 CC transplant rejections and graft versus host disease, infectious diseases
 CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
 CC other blood-related disorders (sickle cell anemia), myeloproliferative
 CC disorders, primary haematopoietic disorders, hyperproliferative disorders
 CC (e.g. Gaucher's disease and cancer), neurodegenerative disorders (e.g.
 CC Alzheimer's disease, Parkinson's disease), chromosomal abnormalities
 CC (Down syndrome), ischemic injury (e.g. stroke), renal disorders (e.g.
 CC glomerulonephritis), cardiovascular disorders (e.g. arrhythmia),
 CC respiratory disorders, dermatological disorders (e.g. wound healing,
 CC epithelial cell proliferation, endocrine disorders (e.g. Addison's
 CC disease), reproductive system disorders, gastrointestinal disorder
 CC (inflammatory disorders), liver disorders (cirrhosis), as stimulators of
 CC B-cell responsiveness to pathogens, activators of T-cells, to induce
 CC higher affinity antibodies, and as a means to induce tumour proliferation
 CC in pathologies e.g. acquired immune deficiency syndrome (AIDS). MAS26976-
 CC MAS27850 represent novel signal transduction pathway protein coding
 CC sequences and PCR primers of the invention

Query Match 14.7%; Score 272.4; DB 4; Length 1966;

Best Local Similarity 56.5%; Pred. No. 1.3e-44; Mismatches 471; Indels 42; Gaps 7;

Matches 667; Conservative 0; Mismatches 471; Indels 42; Gaps 7;

QY 497 GAGAGAAGCTGACGGGCTGGGACGAGATGAGAAGATGAGAGAGTGAAGGAAACC 556
 DB 516 GAGACCCGGGAACTCTCTCAGCGAGATGACGAGAGAGGAGGAGAAACCGGAGGGC 575
 QY 557 TGCCCAAGGCTC---CAAGAGGACCTGAATTGAAAGCTCCAGTGGCAGTGTCCATG 613
 DB 576 CCGGAGAGACCGGGCGGGACGGCTCAACATTGACCACTCCAGTGGCAGTGTCCGATG 635
 QY 614 GAGAGAGACCGGCGGAGAGTGAACCTTCACTCTGATGACTTTGACAACAAAGGCAAG 673
 DB 636 GAGAGAGACCGGCGGAGAGTGAACCTTCACTCTGATGACTTTGACAACAAAGGCAAG 695
 QY 674 GTCACCCGAGAGAGATCACCAGCTTGTGACACATCTATAGAGTGTGAGCTCTCT 733
 DB 696 GTCACCCGAGAGAGATCACCAGCTTGTGACACATCTATAGAGTGTGAGAGCTCTCG 755
 QY 734 GTCACCCGAGAGAGATCACCAGCTTGTGACACATCTATAGAGTGTGAGAGCTCTCT 793
 DB 756 GTCACCCGAGAGAGATCACCAGCTTGTGACACATCTATAGAGTGTGAGAGCTCTCT 815
 QY 794 GGCAGCCAGAGAGAGAGAGTCTTGTCAATCAGGCTGACCTGACAGGCGCAAGGCC 853
 DB 816 CCTCCAGAGAGAGAGAGAGTCTTGTCAATCAGGCTGACCTGACAGGCGCAAGGCC 875
 QY 854 CGAGCAGAGACCAAGCCCACTGAGAGCTTGGAGCTGGAGAGAAAGACGAGCCCG 913
 DB 876 AGGATGAGAGGAGTGAAGTGCAGAGAGCAAGGGTGGCTGACAGAGAGTGTGACAC 935
 QY 914 CTCAGGTTCCAGGTTGACAGCCGCTGAGAGAGTGTGCTGCTACACATTTGGTATGAT 973
 DB 936 GTCAGAGAGGCGCCACTACTGACCCCAAGCTCTGAGAGGCGGGGCTCTACTGTGTGAC 995
 QY 974 GAGAACTGAGAGAGAGAAACCACTATAGATCTCCCGGGATAGAAAACCTACACGCTCC 1033
 DB 996 GAGAACTGAGAGAGAGAGAAACCACTATAGATCTCCCGGGATAGAAAACCTACACGCTCC 1055
 QY 1034 CAATTGGGCTGTGCTCCCTTCTGGTGGCCCAAGAGTCAAACTGCCCCCGGACCTCC 1093
 DB 1056 AGATTGGGCTGTGCTCCCTTCTGGTGGCCCAAGAGTCAAACTGCCCCCGGACCTCC 1115
 QY 1094 AATCCCA---CTGATCTCGTCCCATGAGCGCGGAAGCCATCCCATCCACAGGAAAG 1150
 DB 1116 CACCTCAGAGCGCCGCTCCGCTCCAGAGAGCAATACATGCGGTACACACCGGAGG 1175

QY 1151 CCCAAGGCTGAGACCGGCGCTCTCTCACTCTTGTGACACCCCAATGCGCAAGTCTCA 1210
 DB 1176 TCACAGTGTGTGTGAAACAGCTGTGTGTCAGCTCGA---GCTGTGTGCGGGGCTTG 1232
 QY 1211 GAGCTCAGCAAGGCTCTCGGGGACCCAGAGAGGAGCAAGCACTTTGTAGTCTCC 1270
 DB 1233 GACACGAGCGCCCGCGCGGAGGAGGCGCGAG-----AAGCAATTCTCAAGTCTCC 1283
 QY 1271 AAGGCGCGGAGAGAGT 1330
 DB 1284 AAGGCTCTCGGAGAGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1343
 QY 1331 CCTGTGGAACCGGACATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1390
 DB 1344 TACTACCTGTGCGGCGGTCTC-----TGGCGCCCAAGGCGCTTCAAGAGGCGCACCTTC 1397
 QY 1391 CT 1450
 DB 1398 CCGGAGCGCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1457
 QY 1451 CAGAGAGGCTGCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1510
 DB 1458 CACTGCGCACTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1505
 QY 1511 GAGCAGCTGCGGAGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1570
 DB 1506 GAGGT 1565
 QY 1571 CAGAGCATGAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1624
 DB 1566 CAGGCGCAGAGAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1625
 QY 1625 TACGAGCATGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1664
 DB 1626 CACCGTCTTACGCGCACTGCGCAACGACACTGCTCCCA 1665

RESULT 15
 ADB93295
 ID ADB93295 standard; cDNA; 1966 BP.
 AC ADB93295;
 XX 04-DEC-2003 (first entry)
 DE Human cDNA encoding a novel protein #142.
 XX ss; gene; human; autoimmune disease; Parkinson's disease; silicosis;
 KW gastrointestinal disease; atherosclerosis; haemophilia; thrombocytopenia;
 KW immunosuppressive agent; adjuvant; enhance immune response;
 KW higher affinity antibody induction;
 KW increased serum immunoglobulin concentration.
 XX Homo sapiens.
 OS US2002168711-A1.
 PN US2002168711-A1.
 XX 14-NOV-2002.
 PD 17-JAN-2001; 2001US-0076468.
 PF 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218280P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 22-AUG-2000; 2000US-0226688P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236337P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 08-DEC-2000; 2000US-0251866P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.

(ROSE/) ROSEN C A.
 (RUBEN/) RUBEN S M.
 (BARA/) BARASH S C.

Rosen CA, Ruben SM, Barash SC;

WPI; 2003-719985/68.

P-PSDB; ADB93908.

PT New isolated polypeptide useful for diagnosing and treating
 PT immunosuppressive conditions such as autoimmune disease and Parkinson's
 PT disease.

PS Claim 3; SEQ ID NO 152; 345bp; English.

XX The invention relates to an isolated polypeptide. The polypeptide is
 CC useful for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition in a subject, by determining the presence or
 CC amount of expression of the polypeptide in a biological sample and
 CC diagnosing a pathological condition or a susceptibility to a pathological
 CC condition based on the presence or amount of expression of the
 CC polypeptide. The polypeptide is also useful for identifying a binding
 CC partner to the polypeptide, which involves contacting the polypeptide
 CC with a binding partner and determining whether the binding partner
 CC effects an activity of the polypeptide. The polypeptide or the nucleic
 CC acid encoding the polypeptide is useful for preventing, treating, or
 CC ameliorating a medical condition, which involves administering the
 CC polypeptide or the nucleic acid to a mammalian subject. The nucleic acid
 CC is useful for diagnosing a pathological condition or a susceptibility to
 CC a pathological condition in a subject, which involves determining the
 CC presence or absence of a mutation in the nucleic acid, and diagnosing a
 CC pathological condition or susceptibility to a pathological condition
 CC based on the presence or absence of the mutation. The polypeptide, the
 CC nucleic acid and an antibody to the polypeptide are useful for treating

CC autoimmune disease, Parkinson's disease, silicosis, gastrointestinal
 CC disease, atherosclerosis, haemophilia, thrombocytopenia. The polypeptide,
 CC the nucleic acid and the antibody are useful as immunosuppressive agents,
 CC as adjuvants to enhance immune responses, and as agents to induce higher
 CC affinity antibodies and increase serum immunoglobulin concentrations. The
 CC present sequence represents cDNA encoding a novel human protein. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format direct from USPTO at
 CC seqdata.uspto.gov/sequence.html?docid=20020168711.

XX Sequence 1966 BP; 391 A; 726 C; 608 G; 241 T; 0 U; 0 Other;

Query Match 14.7%; Score 272.4; DB 10; Length 1966;
 Best Local Similarity 56.5%; Pred. No. 1.3e-44;
 Matches 667; Conservative 0; Mismatches 471; Indels 42; Gaps 7;

Qy 497 GAGAACTGACCGGGCTGGGCGGAGATGAGAAAGATGAGAGAGTGGCAACC 556
 Db 516 GAGCACTCCGGGAGCACTCTCAGCGCATGACGAGAGAGGGCAGAAACCGGAGGCC 575
 Qy 557 TGCCAGGCTC---CAAGAGCAGCTGAAGTTGAAGAGCTCCAGTGCATG 613
 Db 576 CCGGAGGAGCCGGGCGGGGAGCGCCCTCAATTGAGCACTCCAGTGCATGTCGTG 635
 Qy 614 GAGAGGAGCAGCCGGGAGGAGTGAACCTTATATATCTTGAACAACAGGCAAG 673
 Db 636 GAGAGGAGCAGCCGGGAGGAGTGAACCTTATATATCTTGAACAACAGGCAAG 695
 Qy 674 GTCACCGGAGGAGCATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 733
 Db 696 GTCACCGGAGGAGCATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 755
 Qy 734 GTCACCACTCCCAATCCAGCAAGATGCTGCGGTAAGTCAACCGGCCCGAT 793
 Db 756 GTCACCACTCTCGGGAGCAGCAAGACCTCGGTGAAGTCAACCGGCCCGAT 815
 Qy 794 GGCAGCCAGAGCAGAGAGAGCTCTTGTCAATCAAGCTACCTGCGAGCGCAAGCCC 853
 Db 816 CCTCCAGCAAGAGGAGAGAGGCTCTCTGCTGCGGCAAGAGCCGAGCCCGTTGC 875
 Qy 854 CGAGCAGAGCAGCAAGCCCACTGAGGAGCTGCGGAGCTGGAGAAAGACGAGCCCG 913
 Db 876 AGATGAGGAGGTAACTGCGAGAGAGCAGAGGAGGTGTCTGCAAC 935
 Qy 914 CTCAGTTCCAGGAGTGAACAGCCGCTGAGAGAGTGTGCTTACCACTTGTGAT 973
 Db 936 GTCAGAGGCGCCAGTACTGACCCCAAGCCCTGCTCGAGAGCGGGGCGCTTGTG 995
 Qy 974 GAGAACTGAGAGAGAGAAACCACTTATGATCTGCGGGATGAGAAACCTACAGTCC 1033
 Db 996 GAGAACTGAGAGAGAGAAACCACTTATGATCTGCGGGATGAGAAACCTACAGTCC 1055
 Qy 1034 CAATTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1093
 Db 1056 AGATTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1115
 Qy 1094 AATTCGA---CTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1150
 Db 1116 CACTTCAGGCGCGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1175
 Qy 1151 CCCCAGGCTGAGACCGGCTCTTCTTCACTTCTTGAACCCCAATTCGCAAGTCTCA 1210
 Db 1176 TCAAGGTGTGTGAGAACATGTCGTCAGCTCGA---GGCTGTGCTGCGGCGCTG 1232
 Qy 1211 GACCTCAGCAAGGCTCGGCGGACCCAGAGAGCGGAGCAAGCACTTGTGAGGTCCCG 1270
 Db 1233 GACACGAGCCCGGCGGAGAGGAGCGGAGC-----AGAGCTTCTCAAGTCCCG 1283
 Qy 1271 AAGGCCAGGAGCAAGGT 1330
 Db 1284 AAGGGCTCGGAGAGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1343
 Qy 1331 CTTCTGGAACCGGCATCTCTGCGGTCTCCCTCGGCGCACTGTGCTGCAAGCCGCG 1390

```
Db      1344 TACTACCTGCGGCGCCGTC-----TGCCGCCCCAGGCCCCCTTCAGGACGGCACCACTTC 1397
QY      1391 CTCTCTCCCTTCCTTAGCCCCCTTGGGCACAAGACCAAGCACCGAGCCAGAGAGC 1450
Db      1398 CCGAGCCCCCACCACCGCTACGCTACAGGCGTACCGCAAAAGGCGAGGAGGC 1457
QY      1451 CAGCAGGGCTGCCGGGGCTGCAGAGCACTGGCTCAGGTGGCCCTGTCTTGGGCGG 1510
Db      1458 CACTCGCCACTCAAGGCCCTCCACAGCTCAGCTTGCACAG-----TGAGCAC 1505
QY      1511 GAGCACCCTGCGGAGCTGCCCGCCTTGTGTATGAGACCAAGCCGGGACCGGTC 1570
Db      1506 GAGGTGGTGGGGAGCTGCCGCCACGCCAGAGAGGGGCTACGCGTGCAGTGATC 1565
QY      1571 CAGAGCATGAGCACCAACCAACCATGAACATACCACTT-----ACACCACTTC 1624
Db      1566 CAGCGGACAGCACCAACCAACCAAGAGCACCAACCAACCAACCACTTC 1625
QY      1625 TACCAGCATAGAGCCCTCCGCCAGGGCCCAACCCTGCA 1664
Db      1626 CACCCGTCTAGCGCCACTGCAAGCAACCTCGCTCCA 1665
```

Search completed: December 29, 2004, 20:56:14
Job time : 1312 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2004, 13:06:18 ; Search time 5335 Seconds
(without alignments)
1968.155 Million cell updates/sec

Title: US-09-993-966-5

Perfect score: 1859

Sequence: 1 gaattccgcccttctataacg.....aaaaaaaaaagcgcgccgc 1859

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4176236 seqs, 2824127955 residues

Total number of hits satisfying chosen parameters: 8352472

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*

16: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

17: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

18: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*

20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1859	100.0	1859	10	US-09-993-966-5
2	1786	96.1	1786	10	US-09-993-966-1
3	1446.4	77.8	1448	13	US-10-087-192-179
4	1130.8	60.8	2807	13	US-10-087-192-176
5	1076.8	57.9	1416	10	US-09-993-966-2
6	1066.6	57.4	1401	9	US-09-730-989-1
7	1066.6	57.4	1401	10	US-09-993-966-6
8	783	42.1	2379	10	US-09-764-891-5902
9	783	42.1	2379	10	US-09-764-891-5904
10	783	42.1	106315	13	US-10-087-192-178
11	590	31.7	590	10	US-09-993-966-21
12	557.2	30.0	98474	13	US-10-087-192-175

13	418.2	22.5	427	10	US-09-918-995-8686	Sequence 8686, App
14	272.4	14.7	1966	9	US-09-764-868-152	Sequence 152, App
15	215	11.6	215	10	US-09-993-966-22	Sequence 22, App
16	207	11.1	1959	18	US-10-723-860-5220	Sequence 5220, App
17	148	8.0	148	10	US-09-993-966-18	Sequence 18, App
18	137	7.4	598	15	US-10-029-386-7990	Sequence 7990, App
19	137	7.4	1950	13	US-10-016-157A-126	Sequence 126, App
20	134	7.2	134	10	US-09-993-966-14	Sequence 14, App
21	133	7.2	133	15	US-10-029-386-21690	Sequence 21690, App
22	128	6.9	128	10	US-09-993-966-20	Sequence 20, App
23	120	6.5	120	10	US-09-993-966-10	Sequence 10, App
24	109	5.9	639	15	US-10-243-552-231	Sequence 231, App
25	107	5.8	107	10	US-09-993-966-15	Sequence 15, App
26	96	5.2	96	10	US-09-993-966-17	Sequence 17, App
27	89	4.8	554	15	US-10-029-386-7615	Sequence 7615, App
28	86.4	4.6	179	15	US-10-029-386-21315	Sequence 21315, App
29	85	4.6	85	10	US-09-993-966-19	Sequence 19, App
30	75	4.0	901	13	US-10-027-632-162722	Sequence 162722, App
31	75	4.0	1817	9	US-09-833-381-1215	Sequence 1215, App
32	72.4	3.9	566	13	US-09-822-849A-83	Sequence 83, App
33	68.2	3.7	566	13	US-10-027-632-266931	Sequence 266931, App
34	68.2	3.7	566	15	US-10-027-632-266931	Sequence 266931, App
35	68.2	3.7	2724	9	US-09-833-381-1215	Sequence 1215, App
36	68.2	3.7	2724	9	US-09-833-381-1215	Sequence 1215, App
37	67	3.6	67	10	US-09-993-966-15	Sequence 15, App
38	65.8	3.5	73038	17	US-10-322-696-37	Sequence 37, App
39	65.6	3.5	670	16	US-10-275-287-19	Sequence 19, App
40	59	3.2	498	9	US-09-783-590-676	Sequence 676, App
41	58.8	3.2	707	15	US-10-072-602B-214	Sequence 214, App
42	57.6	3.1	51657	16	US-10-057-475B-10475	Sequence 10475, App
43	57.6	3.1	51657	16	US-10-154-884B-10475	Sequence 10475, App
44	57.4	3.1	510	16	US-10-275-287-17	Sequence 17, App
45	57.4	3.1	417	16	US-10-275-287-28	Sequence 28, App

ALIGNMENTS

RESULT 1

US-09-993-966-5

Sequence 5, Application US/0993966

Publication No. US20030186232A1

GENERAL INFORMATION:

APPLICANT: ROHAN, MICHAEL

TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF MKD PROTEIN,

FILE REFERENCE: 014024/0280733

CURRENT APPLICATION NUMBER: US/09/993,966

CURRENT FILING DATE: 2001-11-27

PRIOR APPLICATION NUMBER: 60/252,884

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/291,109

PRIOR FILING DATE: 2001-05-16

PRIOR APPLICATION NUMBER: 60/325,571

PRIOR FILING DATE: 2001-10-01

NUMBER OF SEQ ID NOS: 26

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 5

LENGTH: 1859

TYPE: DNA

ORGANISM: Homo sapiens

US-09-993-966-5

Query Match 100.0%; Score 1859; DB 10; Length 1859;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1859; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTGCCCTTTATATGACCTCATATAGGCAAGCACTGTATACACGCGAGTACG 60

DB 1 GAATTGCCCTTTATATGACCTCATATAGGCAAGCACTGTATACACGCGAGTACG 60

QY 61 CGGAGAGTGGGGCGGCGGCGAGCGAGCGAGCGGTCCCGGCGCGCTCGGCTCCG 120

Db	61	CGGGAGATCGGGGCGCGGGCGACCGGCGGAGGAGCGCGTCCGCGGCGCGCTTCGGGACTCCG	120
Qy	121	CTCGGCTTCGGGGGGCTGCTTCGGGAGAGAGAGACCTCAAGGAGGGGCGACGGCCCGCGGGCC	180
Db	121	CTCGGCTTCGGGGGGCTGCTTCGGGAGAGAGAGACCTCAAGGAGGGGCGACGGCCCGCGGGCC	180
Qy	181	GGGCGCATATGGCTTAGGGAAGCTCTCCCGGCGCGGAGCCCGACATATGGGGAAATCTTCACTC	240
Db	181	GGGCGCATATGGCTTAGGGAAGCTCTCCCGGCGCGGAGCCCGACATATGGGGAAATCTTCACTC	240
Qy	241	CAAGCCGGCGCGCGTGTGCAAGCGCAGGGAGAGCCCGGAAGGTGACAGCTTTCGCGGTGAG	300
Db	241	CAAGCCGGCGCGCGTGTGCAAGCGCAGGGAGAGCCCGGAAGGTGACAGCTTTCGCGGTGAG	300
Qy	301	CGCTGCTCGGGCTTCGGAAAGGCGCATTCAGAGATGTGATTCGGAGACAGCGCTTCGCGGGGGG	360
Db	301	CGCTGCTCGGGCTTCGGAAAGGCGCATTCAGAGATGTGATTCGGAGACAGCGCTTCGCGGGGGG	360
Qy	361	TGTCTCGGGAGCCCGGACAGCTGCGGGTTGGCGGGGACCACTAGGCGGAGACA	420
Db	361	TGTCTCGGGAGCCCGGACAGCTGCGGGTTGGCGGGGACCACTAGGCGGAGACA	420
Qy	421	CGTGGGCGACGCTGTTGAGACACGCTTCAGCGAGAGAGAGAGACGACTTTTCGGCTTGA	480
Db	421	CGTGGGCGACGCTGTTGAGACACGCTTCAGCGAGAGAGAGAGACGACTTTTCGGCTTGA	480
Qy	481	AGTGGCCCTGCTCTCTGAGAGACTGACGGGCTGTGGCAGCGGAGATGAGAGAAATGGA	540
Db	481	AGTGGCCCTGCTCTCTGAGAGACTGACGGGCTGTGGCAGCGGAGATGAGAGAAATGGA	540
Qy	541	GAGAGTGAAGGACCCCTGSCCGAGGCTCCAAAGCAGCTGAAGTTTGAAGAGCTCCAGTG	600
Db	541	GAGAGTGAAGGACCCCTGSCCGAGGCTCCAAAGCAGCTGAAGTTTGAAGAGCTCCAGTG	600
Qy	601	CGACGTGCTCATGAGAGAGAGACAGCGGCGAGAGTGAACCTTTCACCCCTGTATGACTTTGA	660
Db	601	CGACGTGCTCATGAGAGAGAGACAGCGGCGAGAGTGAACCTTTCACCCCTGTATGACTTTGA	660
Qy	661	CAACAACGGCAAGGTCACCCGAGAGAGACATCAACGACTTGTGCGACACATCTATGAGGT	720
Db	661	CAACAACGGCAAGGTCACCCGAGAGAGACATCAACGACTTGTGCGACACATCTATGAGGT	720
Qy	721	GGTGGACTCCCTCTGTCAACCACTCCCGCAACATCCAGCAAGATGCTGCGGGTAAAGCTCAC	780
Db	721	GGTGGACTCCCTCTGTGTCAACCACTCCCGCAACATCCAGCAAGATGCTGCGGGTAAAGCTCAC	780
Qy	781	CGTGGCCCGCCATGGCAGGCGACGAGAGCAAGAGAGCGTCTTGTCAATCAGGCTGACCTTGCA	840
Db	781	CGTGGCCCGCCATGGCAGGCGACGAGAGCAAGAGAGCGTCTTGTCAATCAGGCTGACCTTGCA	840
Qy	841	GAGCGCAAGGCGCCGAGGACAGACCCAGCCCACTGAGGACTTGCAGACTGTGGAGAGAA	900
Db	841	GAGCGCAAGGCGCCGAGGACAGACCCAGCCCACTGAGGACTTGCAGACTGTGGAGAGAA	900
Qy	901	GCAAGGAGCCCGGCTCAGGTTCCAGGGTGAACGCGCCTTGGAGCAGTCTTGCTGTACCA	960
Db	901	GCAAGGAGCCCGGCTCAGGTTCCAGGGTGAACGCGCCTTGGAGCAGTCTTGCTGTACCA	960
Qy	961	CCATTGCGTAGATGAGACATCGAGAGAGAAACCATACTTATGATCTCGCGGGATGGA	1020
Db	961	CCATTGCGTAGATGAGACATCGAGAGAGAAACCATACTTATGATCTCGCGGGATGGA	1020
Qy	1021	AAACTACAGTCCCAATTTGGGGCTGAGCTCCCTTCGCTGCGTGGCCGAGAGTCAAGACTGCG	1080
Db	1021	AAACTACAGTCCCAATTTGGGGCTGAGCTCCCTTCGCTGCGTGGCCGAGAGTCAAGACTGCG	1080
Qy	1081	CCCCCGCACTTCCAATCCCACTCGATCTTCGCTCCCATAGCGCGGAGCATTCACATCCC	1140
Db	1081	CCCCCGCACTTCCAATCCCACTCGATCTTCGCTCCCATAGCGCGGAGCATTCACATCCC	1140
Qy	1141	ACACGGAAAGGCCCAAGGCGTGAACCGGGCGTCTTCCACTTCTTGAACCCCAATGCG	1200
Db	1141	ACACGGAAAGGCCCAAGGCGTGAACCGGGCGTCTTCCACTTCTTGAACCCCAATGCG	1200

Qy	1201	CAAGGTCTCAGAGGTCTCAGAGAA	CGGCTCCGGGGCA	CCCAAGAGCGGGAGCAAGCACTTTGT	1260			
Db	1201	CMAAGTCTCAGAGGTCTCAGAGAA	CGGCTCCGGGGCA	CCCAAGAGCGGGAGCAAGCACTTTGT	1260			
Qy	1261	GAGGTCCCCCAGGCGCCAGAGG	CAAGAGTGTGGGTGTGGG	CCACGTGGCCAGAGGGCGAAG	1320			
Db	1261	GAGGTCCCCCAGGCGCCAGAGG	CAAGAGTGTGGGTGTGGG	CCACGTGGCCAGAGGGCGAAG	1320			
Qy	1321	AAAACAACCCCTCTGGGAA	CCCGCAATCCCTGGGGGTGT	CCCCCTCGCGCCCAACCTGGCTGC	1380			
Db	1321	AAAACAACCCCTCTGGGAA	CCCGCAATCCCTGGGGGTGT	CCCCCTCGCGCCCAACCTGGCTGC	1380			
Qy	1381	CAGCCCGGCGCTCTCCGCT	CCCTAGCCCGCCCTCTGGGCA	CAGAAGAGCAAAAGCA	1440			
Db	1381	CAGCCCGGCGCTCTCTCCGCT	CCCTAGCCCGCCCTCTGGGCA	CAGAAGAGCAAAAGCA	1440			
Qy	1441	CAAGAGAGCCAGCAGG	CTGC	CGGGGGCTGC	CAGAGCACTGGCTCTGAGTGGCCCTGT	1500		
Db	1441	CAAAGAGAGCCAGCAGG	CTGC	CGGGGGCTGC	CAGAGCACTGGCTCTGAGTGGCCCTGT	1500		
Qy	1501	CTGTGGGCGGGAGAC	CTGGGGGAGCTGGCCGCT	TGGTGTATGAGAGCA	GGCCGG	1560		
Db	1501	CTGTGGGCGGGAGAC	CTGGGGGAGCTGGCCGCT	TGGTGTATGAGAGCA	GGCCGG	1560		
Qy	1561	GCAGCCGGCTCAGAGACAT	GAGCA	CCACCA	CCATGAACATCACCA	CCATTACCA	1620	
Db	1561	GCAGCCGGCTCAGAGACAT	GAGCA	CCACCA	CCATGAACATCACCA	CCATTACCA	1620	
Qy	1621	CTTCTACCAAGCAT	ATGAGCCCTCC	CCAGGGCCCA	CCCTGC	CAATAGAAAGAC	CCCGAC	1680
Db	1621	CTTCTACCAAGCAT	ATGAGCCCTCC	CCAGGGCCCA	CCCTGC	CAATAGAAAGAC	CCCGAC	1680
Qy	1681	CCCGAGCA	CCCAAGGGCATTTATTT	CAATTAATTAATTAATTAATTAATTTAT	1740			
Db	1681	CCCGAGCA	CCCAAGGGCATTTATTT	CAATTAATTAATTAATTAATTAATTTAT	1740			
Qy	1741	TAATATTAATTAATTTACT	CCACTAATAATTTAG	CTAGCCCTACATGTAGAAATCTAT	TGAAA	1800		
Db	1741	TAATATTAATTAATTTACT	CCACTAATAATTTAG	CTAGCCCTACATGTAGAAATCTAT	TGAAA	1800		
Qy	1801	CACGAACTAAACCTTTATTTAT	TATGTATAAAAAA	AAAAAAGCGCGC	1859			
Db	1801	CACGAACTAAACCTTTATTTAT	TATGTATAAAAAA	AAAAAAGCGCGC	1859			

RESULT 2

US-09-993-966-1

Sequence 1, Application US/09993966

Publication No. US2003018623A1

GENERAL INFORMATION:

APPLICANT: ROHANI, MICHAEL

TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NKD PROTEIN, TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF

FILE REFERENCE: 014024/028073

CURRENT APPLICATION NUMBER: US/09/993,966

PRIOR FILING DATE: 2001-11-27

PRIOR APPLICATION NUMBER: 60/252,884

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/291,109

PRIOR FILING DATE: 2001-05-16

PRIOR APPLICATION NUMBER: 60/325,571

PRIOR FILING DATE: 2001-10-01

NUMBER OF SEQ ID NOS: 26

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 1786

TYPE: DNA

ORGANISM: Homo sapiens

US-09-993-966-1

Query Match 96.1%; Score 1786; DB 10; Length 1786;

Best Local Similarity 100.0%; Pted No. 0;

Query Match 77.8%; Score 1446.4; DB 13; Length 1448;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1447; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 186 CATGGCTTAAGGAGAGCTCCCGGCGCGGAGCCCGCAGCATGAGGAGAACTTCACTCCAAAGC 245
 Db 1 CATGGCTTAAGGAGAGCTCCCGGCGCGGAGCCCGCAGCATGAGGAGAACTTCACTCCAAAGC 60

QY 246 CGGCGCGCTGTGTGCAAGCGCAGGAGAGCCCGGAAAGGTGACAGCTTGCCTGAGCGCTG 305
 Db 61 CGGCGCGCTGTGTGCAAGCGCAGGAGAGCCCGGAAAGGTGACAGCTTGCCTGAGCGCTG 120

QY 306 CCGTGGCTCGGAGAGGAGATGAGATCGGAGAGCAAGCGCTGCGCGCGGAGCTGCT 365
 Db 121 CCGTGGCTCGGAGAGGAGATGAGATCGGAGAGCAAGCGCTGCGCGCGGAGCTGCT 180

QY 366 CGGAGACCCCGGACAGCTCGGTTGGCGGGCACATAGGCGGAGAGCCCGGAGCTCGTGG 425
 Db 181 CGGAGACCCCGGACAGCTCGGTTGGCGGGCACATAGGCGGAGAGCCCGGAGCTCGTGG 240

QY 426 GCGACGTGTGAGAGACACCTTCAGCGAGAGAGAGAGAGCATTTGCGCTGGAAGTGG 485
 Db 241 GCGACGTGTGAGAGACACCTTCAGCGAGAGAGAGAGAGCATTTGCGCTGGAAGTGG 300

QY 486 CCGTGGCTCGGAGAGGAGCTGAGCGGGCTGGCAGCGGAGATGAGAGAGATGAGAGAG 545
 Db 301 CCGTGGCTCGGAGAGGAGCTGAGCGGGCTGGCAGCGGAGATGAGAGAGATGAGAGAG 360

QY 546 TGAAGCGAACCTGCGCGAGGCTCCAAAGAGCAGTGAAGTTGAAGAGCTCCAGTGCAGC 605
 Db 361 TGAAGCGAACCTGCGCGAGGCTCCAAAGAGCAGTGAAGTTGAAGAGCTCCAGTGCAGC 420

QY 606 TGTTCATGAGAGAGAGAGAGCGCGGAGAGTGAAGCTTCACTCTGTATGACTTTGACAA 665
 Db 421 TGTTCATGAGAGAGAGAGAGCGCGGAGAGTGAAGCTTCACTCTGTATGACTTTGACAA 480

QY 666 ACCGCAAGGTCAACCCGAGAGGACATCAACAGCTTGTGCACACCATCATAGAGGTGG 725
 Db 481 ACCGCAAGGTCAACCCGAGAGGACATCAACAGCTTGTGCACACCATCATAGAGGTGG 540

QY 726 ACTCTCTGTCAACCACTCCCAACATTCAGCAAGATGCTGGGGTAAAGCTCAACGCTG 785
 Db 541 ACTCTCTGTCAACCACTCCCAACATTCAGCAAGATGCTGGGGTAAAGCTCAACGCTG 600

QY 786 CCCCCGATGGGAGCCAGAGCAAGAGAGCTTGTTCATCAAGCTGACCTGCAGAGCG 845
 Db 601 CCCCCGATGGGAGCCAGAGCAAGAGAGCTTGTTCATCAAGCTGACCTGCAGAGCG 660

QY 846 CAAGGCCCCGAGGAGAGCCAAAGCCCACTGAGGACCTGGGAGCTGGGAGAGAAAGAGC 905
 Db 661 CAAGGCCCCGAGGAGAGCCAAAGCCCACTGAGGACCTGGGAGCTGGGAGAGAAAGAGC 720

QY 906 GAGCGCCGCTCAGGTTCCAGGGTGAAGCCGCTGAGACAGTCTGGCTGTAACCACT 965
 Db 721 GAGCGCCGCTCAGGTTCCAGGGTGAAGCCGCTGAGACAGTCTGGCTGTAACCACT 780

QY 966 GCGTATGATGAGAAATGAGAGAGAAACCACTACTTATGATCTGCGGGGATAGAAAAT 1025
 Db 781 GCGTATGATGAGAAATGAGAGAGAAACCACTACTTATGATCTGCGGGGATAGAAAAT 840

QY 1026 ACAGCTCCCAATTTGGGCTGGCTCCCGCTTCCGTTGGGCGCAGAAATGAGAACTGGCCCC 1085
 Db 841 ACAGCTCCCAATTTGGGCTGGCTCCCGCTTCCGTTGGGCGCAGAAATGAGAACTGGCCCC 900

QY 1086 GCACCTCCCAATCCCACTGATCTGCTCCATGAGCCGGAAGCCATCCATCCCAAGC 1145
 Db 901 GCACCTCCCAATCCCACTGATCTGCTCCATGAGCCGGAAGCCATCCATCCCAAGC 960

QY 1146 GAAAGCCCCCAAGGGGTGAGACCCGGCTCTCTTCACTTCTTGAACAACCCCAATCCGCAAG 1205
 Db 961 GAAAGCCCCCAAGGGGTGAGACCCGGCTCTCTTCACTTCTTGAACAACCCCAATCCGCAAG 1020

QY 1206 TCTCAGAGCTCCAGCAACGGCTCCGGGGACACCGAGAGCGGAGCAACATTTGTAGAGT 1265
 Db 1021 TCTCAGAGCTCCAGCAACGGCTCCGGGGACACCGAGAGCGGAGCAACATTTGTAGAGT 1080

QY 1266 CCCCCAAGGCCCAAGGCAAGAGTGTGGGTGGGCACTGGGCAAGGGGCAAGAAACA 1325
 Db 1081 CCCCCAAGGCCCAAGGCAAGAGTGTGGGTGGGCACTGGGCAAGGGGCAAGAAACA 1140

QY 1326 AGCCCCCTGTGGAGACCCGCCATCTGCGGTGTCCCCCTCCGGCCACTGGTGGCCAGCC 1385
 Db 1141 AGCCCCCTGTGGAGACCCGCCATCTGCGGTGTCCCCCTCCGGCCACTGGTGGCCAGCC 1200

QY 1386 CGGCGCTCTCCCTCCCTAAGCCCTCGGAGCAAGAGCAAGCAACGAGCCAAAG 1445
 Db 1201 CGGCGCTCTCCCTCCCTAAGCCCTCGGAGCAAGAGCAAGCAACGAGCCAAAG 1260

QY 1446 AGAGCCAGAGGAGCTGCGGGGCTTGAGGACCACTGGCTTCAAGTGGCTGTCTGG 1505
 Db 1261 AGAGCCAGAGGAGCTGCGGGGCTTGAGGACCACTGGCTTCAAGTGGCTGTCTGG 1320

QY 1506 GCGGGAGACACCTGCGGGAGCTGCGGCTTGGTGTATGAGAGCCAGGCGGGCAGC 1565
 Db 1321 GCGGGAGACACCTGCGGGAGCTGCGGCTTGGTGTATGAGAGCCAGGCGGGCAGC 1380

QY 1566 CGGTCCAGAGACATGAGACACACCAACCATGAACTCAACCAATTACCACTTCT 1625
 Db 1381 CGGTCCAGAGACATGAGACACACCAACCATGAACTCAACCAATTACCACTTCT 1440

QY 1626 ACCAGACA 1633
 Db 1441 ACCAGACA 1448

RESULT 4
 US-10-087-192-176
 / Sequence 176, Application US/10087192
 / Publication No. US20020182586A1
 / GENERAL INFORMATION:
 / APPLICANT: Morris, David W.
 / TITLE OR INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
 / FILE REFERENCE: 529452000122
 / CURRENT APPLICATION NUMBER: US/10/087,192
 / CURRENT FILING DATE: 2002-03-01
 / PRIOR APPLICATION NUMBER: US 09/747,377
 / PRIOR FILING DATE: 2000-12-22
 / PRIOR APPLICATION NUMBER: US 09/798,586
 / PRIOR FILING DATE: 2001-03-02
 / NUMBER OF SEQ ID NOS: 2059
 / SOFTWARE: FastSeq for Windows Version 4.0
 / SEQ ID NO 176
 / LENGTH: 2807
 / TYPE: DNA
 / ORGANISM: Mus musculus
 / US-10-087-192-176

Query Match 60.8%; Score 1130.8; DB 13; Length 2807;
 Best Local Similarity 84.6%; Pred. No. 4,9e-260;
 Matches 1311; Conservative 0; Mismatches 222; Indels 17; Gaps 3;

QY 282 GTGACAGCTTGGCGTGAAGGCTGCGGTGGCTGGAAAGGAGATCGAGAGATCGGGA 341
 Db 1 GTGACAGCTTGGCGTGAAGGCTGCGGTGGCTGGAAAGGAGATCGAGAGATCGGGA 60

QY 342 GAGAGCGCTGCCGGGCGGTGTCTGGGACCCGAGACGTGCGGTTGGCGGGACCTAG 401
 Db 61 GAGAGCGCTGCCGGGCGGTGTCTGGGACCCGAGACGTGCGGTTGGCGGGACCTAG 120

QY 402 GCGGAGACACCCGGGAGCTGTGGGCGAGCTGTGAGAGACAGCTTCAAGGAGAAAG 461
 Db 121 GCGGAGACACCCGGGAGCTGTGGGCGAGCTGTGAGAGAGCTTCTGAGAGAGAGC 180

QY 462 AGGACGCTTTGGGCTGGAGTGGCCCTGCTCTGAGAGACTGAGCGGCTGGGCAAGC 521
DB 181 AGGACGCTTTGGGCTGGAGTGGCCCTGCTCTGAGAGACTGAGCGGCTGGGCAAGC 240
QY 522 GAGATGAG 581
DB 241 GAGATGAG 300
QY 582 AGTTGAG 641
DB 301 AGTTGAG 360
QY 642 TCACCTGATGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 701
DB 361 TCACCTGATGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
QY 702 TGACACCATCTATGAGAGTGGTGGAGTCTCTGTCAACCACTCCCAACATCCAGCAGA 761
DB 421 TGACATCATCTATGAGAGTGGTGGAGTCTCTGTCAACCACTCCCAACATCCAGCAGA 480
QY 762 TGCTGGGGGTAAAGCTGACCGTGGCCCGGATGGCAGCAGAGAGAGAGAGAGAGAG 821
DB 481 CACTGGGGGTAAAGCTGACCGTGGCCCGGATGGCAGCAGAGAGAGAGAGAGAGAGAG 540
QY 822 TCATTCAGGCTGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 881
DB 541 TCATTCAGGCTGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 882 TGCGGAGCTGGGAG 941
DB 601 TGCGGAGCTGGGAG 660
QY 942 AGCAGTCTGGCTGTACCACTTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1001
DB 661 AGCAGCTGAGCTGTACCACTTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 1002 TAGATCTGGCGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1061
DB 721 TAGATCTGGCGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 1062 CCCAGAGTGAAG 1121
DB 781 CCCAGAGTGAAG 840
QY 1122 CGGAGAGCATCTCACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1181
DB 841 CAGAGAGTGGCCCATCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 1182 TCCTTGAACATCCCATCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1241
DB 901 TCCTTGAACATCCCATCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
QY 1242 ACCGAG 1301
DB 961 ATGGAG 1020
QY 1302 ACGTGGAG 1361
DB 1021 ACGGAG 1080
QY 1362 CCTTCGAG 1421
DB 1081 CCTTCGAG 1140
QY 1422 AGAAG 1481
DB 1141 AGAAG 1200
QY 1482 TGGGCTGAG 1538
DB 1201 TGGGCTGAG 1260
QY 1539 TGGGCTGAG 1598

DB 1261 TGGGCTGAG 1320
QY 1599 AACATCACACCATTTACACCATTTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1658
DB 1321 AACATCACACCATTTACACCATTTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1369
QY 1659 CTGCGATATGAG 1718
DB 1370 CTGCGAG 1426
QY 1719 TTATTATGATGATTTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1778
DB 1427 TTATTATGATGATTTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1486
QY 1779 ACATGTAGAGAGATCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1828
DB 1487 TCATGTAG 1536

RESULT 5
US-09-993-966-2
Sequence 2, Application US/09993966
Publication No. US20030186232A1
GENERAL INFORMATION:
APPLICANT: ROHAN, MICHAEL
TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NKD PROTEIN,
NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF
FILE REFERENCE: 014024/028073
CURRENT APPLICATION NUMBER: US/09/993,966
CURRENT FILING DATE: 2001-11-27
PRIOR APPLICATION NUMBER: 60/252,884
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/291,109
PRIOR FILING DATE: 2001-05-16
PRIOR APPLICATION NUMBER: 60/325,571
PRIOR FILING DATE: 2001-10-01
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 1416
TYPE: DNA
ORGANISM: Mus sp.
US-09-993-966-2

Query Match 57.9%; Score 1076.8; DB 10; Length 1416;
Best Local Similarity 85.5%; Pred. No. 3,1e-247;
Matches 1211; Conservative 0; Mismatches 202; Indels 3; Gaps 1;
QY 224 ATGGGAGAACTTCACTCAAG 283
DB 1 ATGGGAGAACTTCACTCAAG 60
QY 284 GAGAGCTTGGCGGAG 343
DB 61 GAGAGCTTGGCGGAG 120
QY 344 CAGGCTGCGGAG 403
DB 121 CAGGCTGCTCAAG 180
QY 404 CGAAG 463
DB 181 CGAAG 240
QY 464 GAGAGCTTGGCGGAG 523
DB 241 GAGAGCTTGGCGGAG 300
QY 524 GATGAG 583
DB 301 GATGAG 360

QY 584 TTGAAAGAGCTCAGTGCAGCTGTCCATGAGAGGAGACAGCCGGACAGAGTGGACCTTC 643
Db 361 TTGGAAGAGCTACAGTGTGTGTCTGTGAGAGAGACAGCCGGACAGAGTGGACCTTC 420
QY 644 ACCCTGATGACTTTGACAAACACGAGAGGTCAACCGAGAGACATCAACAGCTTTCG 703
Db 421 ACTCTATATGACTTCGACAACTAAGGACAAAGTGACCCGTGAGACATTTACAGCTTCG 480
QY 704 CACACCATATAGAGGTGTGACTCTCTGTCAACCACTCCCAATCCAGAAATG 763
Db 481 CATACCATATAGAGGTGTGACTCTCTGTGACCAATTCCTCCATCAAGCAACA 540
QY 764 CTGGGCTAAAGCTCAACGCTGACCCCGATGACAGCAAGAGAGAGTCTTGTG 823
Db 541 CTGGGGTGAAGCTCAACGCTGCTCTGAGGAGACAGAGTGAAGAGAGTCTTTC 600
QY 824 AATCAGGCTGACTGCAAGAGCGCAAGGCCCCAGACAGAACCAAGCCACTGAGACCTG 883
Db 601 AACCATACCATCTGCAAGAGCAAGGCCCCAGACAGCAACCAACCCGCTGAGAGCTG 660
QY 884 CGGAGCTGGAG 943
Db 661 CGTGGCTGGAG 720
QY 944 CAGTCTGCTGCTCAACCAATTCGATGAGAGACATCGAGAGAGAGAGAGAGAGAG 1003
Db 721 CAGGCAACATCTGCTCAACCAATTCGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 1004 GATCTGCTGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1063
Db 781 GACTGCTGGAG 840
QY 1064 CAGAGCTGAG 1123
Db 841 CAGAGCTGAG 900
QY 1124 GAGAGCTGAG 1183
Db 901 GAGAGCTGAG 960
QY 1184 CTTGACACCCCATTCGCAAGAGCTCAAGAGCTCAAGAGAGAGAGAGAGAGAGAG 1243
Db 961 CTTGACACCCCATTCGCAAGAGCTCAAGAGCTCAAGAGAGAGAGAGAGAGAGAG 1020
QY 1244 GGGAGCAG 1303
Db 1021 GGGAGCAG 1080
QY 1304 GTGGCAG 1363
Db 1081 GGGCAG 1140
QY 1364 TCCGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1423
Db 1141 TCTCCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
QY 1424 AAGCAG 1483
Db 1201 AAGCAG 1260
QY 1484 GCTCAGAGTGG---CCCTGTCTGTGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1540
Db 1261 GCTCAG 1320
QY 1541 GTGTATGAG 1600
Db 1321 GTGTATGAG 1380
QY 1601 CATCAGCAGATTAACCACTTCTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1636
Db 1381 CATCAGCAGATTAACCACTTCTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1416

RESULT 6
US-09-730-989-1
; Sequence 1, Application US/09730989
; Patent No. US2002006152A1
; GENERAL INFORMATION:
; APPLICANT: Yan, Dong
; APPLICANT: Williams, Lewis T.
; TITLE OF INVENTION: MAMMALIAN DISHEVELLED-ASSOCIATED PROTEINS
; FILE REFERENCE: PP-01657.002 / 200130.518
; CURRENT APPLICATION NUMBER: US/09730.989
; CURRENT FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1401
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-730-989-1

Query Match 57.4%; Score 1066.6; DB 9; Length 1401;
Best Local Similarity 85.6%; Pred. No. 8.4e-245;
Matches 1199; Conservative 0; Mismatches 199; Indels 3; Gaps 1;

QY 224 ATGGGAAACCTTCACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 283
Db 1 ATGGGAAACCTTCACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
QY 284 GACAGCTTCCGCTGAG 343
Db 61 GACAGCTTCCGCTGAG 120
QY 344 CAGGCTGAG 403
Db 121 CAGGCTGAG 180
QY 404 CGAAGCAG 463
Db 181 CGAAGCAG 240
QY 464 GAGCAGCTTGGCTGAG 523
Db 241 GAGCAGCTTGGCTGAG 300
QY 524 GATGAG 583
Db 301 GATGAG 360
QY 584 TTGGAAGAGCTCAAGTGCAGCTGTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 643
Db 361 TTGGAAGAGCTCAAGTGCAGCTGTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
QY 644 ACCCTGATGACTTTGACAAACACGAGAGGTCAACCGAGAGACATCAACAGCTTTCG 703
Db 421 ACTCTATATGACTTCGACAACTAAGGACAAAGTGACCCGTGAGACATTTACAGCTTCG 480
QY 704 CACACCATATAGAGGTGTGACTCTCTGTCAACCACTCCCAATCCAGAAATG 763
Db 481 CATACCATATAGAGGTGTGACTCTCTGTGACCAATTCCTCCATCAAGCAACA 540
QY 764 CTGGGCTAAAGCTCAACGCTGACCCCGATGACAGCAAGAGAGAGAGAGAGAGAGAG 823
Db 541 CTGGGGTGAAGCTCAACGCTGCTCTGAGGAGACAGAGTGAAGAGAGAGAGAGAGAG 600
QY 824 AATCAGGCTGACTGCAAGAGCGCAAGGCCCCAGACAGAACCAAGCCACTGAGACCTG 883
Db 601 AACCATACCATCTGCAAGAGCAAGGCCCCAGACAGCAACCAACCCGCTGAGAGCTG 660
QY 884 CGGAGCTGGAG 943
Db 661 CTGGCTGGAG 720
QY 944 CAGTCTGCTGCTCAACCAATTCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1003

Db 721 CAGCAGACTGCTACACACATGCGTGATGAGAAACATGAGAGAAACCACTACTTA 780
Qy 1004 GATCTGCGCGGATAGAAAATCAACAGTCCCAATTTGGGCTGCTCCCTTCCGTGAGC 1063
Db 781 GACCTGCGGGATAGAGAACTACACGCTCAGTTTGGACCGGATCCCTTCCGTGAGC 840
Qy 1064 CAGAACTCAGAACTGCCCCCCCCGACCTCCCAATCCCACTGATTCCTGCTCCCAAGACCG 1123
Db 841 CAGAACTCAGAACTGCCCCCCCCGACCTCCCAATCCCACTGATTCCTGCTCCCAAGACCG 900
Qy 1124 GAAGCCATCCATCCCAACCAAGAGCCCAAGAGCCGAGCTCCCTTCCACTTC 1183
Db 901 GAAGCTCCCACTCCCAACCAAGAGCCCAAGAGCTCCCTTCCACTTC 960
Qy 1184 CTTGACACCCCAATCCCAAGAGCTCCCAAGAGCTCCCTTCCACTTC 1243
Db 961 CTTGACACCCCAATCCCAAGAGCTCCCAAGAGCTCCCTTCCACTTC 1020
Qy 1244 GGGAGCAAGCACTTTGAGAGTCCCCCAAGAGCTCCCAAGAGCTTTGGGCTAC 1303
Db 1021 GGGAGCAAGCACTTTGAGAGTCCCCCAAGAGCTCCCAAGAGCTTTGGGCTAC 1080
Qy 1304 GTGAGCAAGAGAGCAAGAAACAGAGCCCTCTGAGAGCCGAGCTCCCTGAGTCCGCC 1363
Db 1081 GGGAGCAAGAGAGCAAGAAACAGAGCTCCCAAGAGCTCCCTGAGTCCGCC 1140
Qy 1364 TCCGCGCACTGAGTCCCAAGAGCCCTCTGAGAGCCGAGCTCCCTGAGTCCGCC 1423
Db 1141 TCCGCGCACTGAGTCCCAAGAGCCCTCTGAGAGCCGAGCTCCCTGAGTCCGCC 1200
Qy 1424 AAGCAAGCAAGCAAGAGCAAGAGAGCAAGAGAGTCCGAGAGCTCCCAAGAGCACTG 1483
Db 1201 AAGCAAGCAAGCAAGAGCAAGAGAGCAAGAGAGTCCGAGAGCTCCCAAGAGCACTG 1260
Qy 1484 GCTCAGAGTGA---CCCTGCTCTGAGAGAGAGAGCAAGAGTCCGAGAGCTCCGAGTGA 1540
Db 1261 GCTCAGAGAGAGTCCCAAG 1380
Qy 1541 GTGATGAG 1600
Db 1321 GTGATGAG 1380
Qy 1601 CATCACCACCATTTACACAC 1621
Db 1381 CATCACCACCATTTACACAC 1401

RESULT 7
US-09-993-966-6
; Sequence 6, Application US/09993966
; Publication No. US20030186232A1
; GENERAL INFORMATION:
; APPLICANT: ROHAN, MICHAEL
; TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NMD PROTEIN,
; FILE REFERENCE: 014024/028073
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: US/09/993,966
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/252,884
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/325,571
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1401
; TYPE: DNA
; ORGANISM: Mus sp.
US-09-993-966-6

Query Match 57.4%; Score 1066.6; DB 10; Length 1401;

Best Local Similarity 85.6%; Pred. No. 84e-245;
Matches 1199; Conservative 0; Mismatches 199; Indels 3; Gaps 1;
Qy 224 ATGGGGAATCTTCACTCAAG 283
Db 1 ATGGGGAATCTTCACTCAAG 60
Qy 284 GACAGCTTCCCTGAG 343
Db 61 GACAGCTTCCCTGAG 120
Qy 344 CAGAGCTTCCCTGAG 403
Db 121 CAGAGCTTCCCTGAG 180
Qy 404 CAGAGCTTCCCTGAG 463
Db 181 CAGAGCTTCCCTGAG 240
Qy 464 GACAGCTTCCCTGAG 523
Db 241 GACAGCTTCCCTGAG 300
Qy 524 GATGAG 583
Db 301 GATGAG 360
Qy 584 TTTGAG 643
Db 361 TTTGAG 420
Qy 644 ACCCTGATGATCTTGAACAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 703
Db 421 ACTGATGATGATCTTGAACAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Qy 704 CACACCATCTATGAG 763
Db 481 CATACCATCTATGAG 540
Qy 764 CTGCGGATGATGATCTTGAACAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 823
Db 541 CTGCGGATGATGATCTTGAACAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Qy 824 AATCAG 883
Db 601 AACCATCTGATGATGATCTTGAACAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Qy 884 CGAG 943
Db 661 CGAG 720
Qy 944 CAGCTGAG 1003
Db 721 CAGCTGAG 780
Qy 1004 GATCTGCGCGGATAGAAAATCAACAGTCCCAATTTGGGCTGCTCCCTTCCGTGAGC 1063
Db 781 GACCTGCGGGATAGAGAACTACACGCTCAGTTTGGACCGGATCCCTTCCGTGAGC 840
Qy 1064 CAGAACTCAGAACTGCCCCCCCCGACCTCCCAATCCCACTGATTCCTGCTCCCAAGACCG 1123
Db 841 CAGAACTCAGAACTGCCCCCCCCGACCTCCCAATCCCACTGATTCCTGCTCCCAAGACCG 900
Qy 1124 GAAGCCATCCATCCCAACCAAGAGCCCAAGAGCTCCCTTCCACTTC 1183
Db 901 GAAGCTCCCACTCCCAACCAAGAGCCCAAGAGCTCCCTTCCACTTC 960
Qy 1184 CTTGACACCCCAATCCCAAGAGCTCCCAAGAGCTCCCTTCCACTTC 1243
Db 961 CTTGACACCCCAATCCCAAGAGCTCCCAAGAGCTCCCTTCCACTTC 1020
Qy 1244 GGGAGCAAGCACTTTGAGAGTCCCCCAAGAGCTCCCAAGAGCTTTGGGCTAC 1303

```
Db 1021 GGGAGCAAGCACTTGTGAGGTGTCCTCCCAAGGCCAAGGGCAAGAAATGAGGTATGGGCGAC 1080
Qy 1304 GTGGCCAGAGGGGCAAGAAAGAGAGCCCTCTGTGAGACCCGCAATCCCTGCGGTGTCCCCC 1333
Db 1081 GGGGCCAGAGGTGCAAGAAAGCAAGCTTCACACTGTGTACCAACCACTACTGTCTCCCCC 1140
Qy 1364 TCCGCCACCTGTGGCAGGCCGCGCTCTCCCTCCCTGAGGCCCCCTCGGAGCAAG 1423
Db 1141 TCTGCCATCTGGCCAGCCAGCCCTCTTCTCCCACTCTGGACCTCTGGGAGCAAG 1200
Qy 1424 AAGCAAGCAAGCCAGGCCAAGAGAGCCAGAGGCTGCGGGGCTTGACAGCACTG 1483
Db 1201 AAACACAAGATGAGGCAAGAGAGCCAGGCTGCGGGGCTTGACAGGCCCCCTG 1260
Qy 1484 GCTCTAGGTG---CCCTGTCTCTGGGGGAGAGACCTGCGGGAGCTGCGGCTTGGTG 1540
Db 1261 GCTGCAAGAGGCTTCACCTGTATGGGGGAGAGAGGTGAGGGAGCTGCGGCTGGTG 1320
Qy 1541 GTGTATGAGAGCCAGGCCGAGGCCGAGTCCAGAGACATGAGCACCAACCACTATGA 1600
Db 1321 GTGTACAGAGCCAGGCTGGGAGAGCCGTCCAGAGACACAGAACCACTACCAACCA 1380
Qy 1601 CATCACCACTATACCAAC 1621
Db 1381 CATCACCACTATACCAAC 1401

RESULT 8
US-09-764-891-5902
; Sequence 5902, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764, 891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5902
; LENGTH: 2379
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-5902

Query Match 42.1%; Score 783; DB 10; Length 2379;
Best Local Similarity 100.0%; Pred. No. 7.9e-177;
Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1046 GGCCTCCCTTCCTGCGCCCAAGATCAGAACTGCCCCCGCACCTTCAATCCCACTCGA 1105
Db 840 GGCCTCCCTTCCTGCGCCCAAGATCAGAACTGCCCCCGCACCTTCAATCCCACTCGA 899
Qy 1106 TCTGCTCCCATGAGCGGAGAGCATTCATCCACACGAAAGCCCAAGGCGTGGAC 1165
Db 900 TCTGCTCCCATGAGCGGAGAGCATTCATCCACACGAAAGCCCAAGGCGTGGAC 959
Qy 1166 CCGGCTCTCTTCACTTCTTGACACCCCAATGCGCAAGTCTCAGAGCTTCAGCAACGG 1225
Db 960 CCGGCTCTCTTCACTTCTTGACACCCCAATGCGCAAGTCTCAGAGCTTCAGCAACGG 1019
Qy 1226 CTCCGGGAGCAAGAGAGAGAGCACTTTGTGAGTCTCCCAAGGCCAAGGCGCAAG 1285
Db 1020 CTCCGGGAGCAAGAGAGAGAGCACTTTGTGAGTCTCCCAAGGCCAAGGCGCAAG 1079
Qy 1286 AGTGTGAGTGTGGAGCAAGTGGCCAGAGGGGCAAGAAACAAGCCCTCTGAGACCGGCC 1345
Db 1080 AGTGTGAGTGTGGAGCAAGTGGCCAGAGGGGCAAGAAACAAGCCCTCTGAGACCGGCC 1139
Qy 1346 ATCCCTGAGTGTCCCTCTCCGCCCACTTGACTGCGAGGCCGCTCTCTCCCTCTCTA 1405
Db 1140 ATCCCTGAGTGTCCCTCTCCGCCCACTTGACTGCGAGGCCGCTCTCTCCCTCTCTA 1199
```

```
Qy 1406 GCCCCCTGGGAGCAAGAGCAAGAGCAAGAGCCAGAGGAGAGGCAAGGAGGCTGCGG 1465
Db 1200 GCCCCCTGGGAGCAAGAGCAAGAGCAAGAGCCAGAGGAGAGGCAAGGAGGCTGCGG 1259
Qy 1466 GGCCTGAGAGCAACCACTGCTCAGGTGCGCTGTCTGTGGGGGAGAGCACTTGCGGAG 1525
Db 1260 GGCCTGAGAGCAACCACTGCTCAGGTGCGCTGTCTGTGGGGGAGAGCACTTGCGGAG 1319
Qy 1526 CTGCGGCTTGTGTGTATGAGAGCAGGCGGAGAGCCGCTTCCAGAGACATGAGAC 1585
Db 1320 CTGCGGCTTGTGTGTATGAGAGCAGGCGGAGAGCCGCTTCCAGAGACATGAGAC 1379
Qy 1586 CACCAACCAATGAACATGACCAACATACACCACTTACACAGACATAGAGCCCTCC 1645
Db 1380 CACCAACCAATGAACATGACCAACATACACCACTTACACAGACATAGAGCCCTCC 1439
Qy 1646 CAGAGGCCCAACCTGCAATATGAGAGCCCAACCCCGACACACCAAGGCAATATAT 1705
Db 1440 CAGAGGCCCAACCTGCAATATGAGAGCCCAACCCCGACACACCAAGGCAATATAT 1499
Qy 1706 CTATTAATTATGTATATGATGATATTTATTAATTAATTAATTTACTCACTAAT 1765
Db 1500 CTATTAATTATGTATATGATGATATTTATTAATTAATTAATTTACTCACTAAT 1559
Qy 1766 ATTTAGCTAGCTTAACATGTAAGATCTATGAAACACAGACCTAACTTTATTTAT 1825
Db 1560 ATTTAGCTAGCTTAACATGTAAGATCTATGAAACACAGACCTAACTTTATTTAT 1619
Qy 1826 GTT 1828
Db 1620 GTT 1622
```

```
RESULT 9
US-09-764-891-5904
; Sequence 5904, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764, 891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5904
; LENGTH: 2379
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-5904

Query Match 42.1%; Score 783; DB 10; Length 2379;
Best Local Similarity 100.0%; Pred. No. 7.9e-177;
Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1046 GGCCTCCCTTCCTGCGCCCAAGATCAGAACTGCCCCCGCACCTTCAATCCCACTCGA 1105
Db 840 GGCCTCCCTTCCTGCGCCCAAGATCAGAACTGCCCCCGCACCTTCAATCCCACTCGA 899
Qy 1106 TCTGCTCCCATGAGCGGAGAGCATTCATCCACACGAAAGCCCAAGGCGTGGAC 1165
Db 900 TCTGCTCCCATGAGCGGAGAGCATTCATCCACACGAAAGCCCAAGGCGTGGAC 959
Qy 1166 CCGGCTCTCTTCACTTCTTGACACCCCAATGCGCAAGTCTCAGAGCTTCAGCAACGG 1225
Db 960 CCGGCTCTCTTCACTTCTTGACACCCCAATGCGCAAGTCTCAGAGCTTCAGCAACGG 1019
Qy 1226 CTCCGGGAGCAAGAGAGAGAGCACTTTGTGAGTCTCCCAAGGCCAAGGCGCAAG 1285
Db 1020 CTCCGGGAGCAAGAGAGAGAGCACTTTGTGAGTCTCCCAAGGCCAAGGCGCAAG 1079
```

OY		1286	AGTGTGGATGTTGGCCACAGTGTCGCCACAGAGGCGCAAGAACCCTCTGTGGACCCGCC	1345
Db		1080	AGTGTGGATGTTGGCCACAGTGTCGCCACAGAGGCGCAAGAACCCTCTGTGGACCCGCC	1139
OY		1346	ATCCTCTGGGTGTCCCCCTCCGCCAACCTGGCTGCAGCCCGGCTCTCTCCCTCCCTA	1405
Db		1140	ATCCTCTGGGTGTCCCCCTCCGCCAACCTGGCTGCAGCCCGGCTCTCTCCCTCCCTA	1199
OY		1406	GCCCCCCTCGGGGACAAGAAGCACAGACCGGACCGAAGAGACCGACAGGCTGCGCG	1465
Db		1200	GCCCCCCTCGGGGACAAGAAGCACAGACCGGACCGAAGAGACCGAAGAGCTGCGCG	1259
OY		1466	GGCCTGACAGGACCACTGGCTCAGGTGCCCTGTCTCTGGGCGGAGCACCTGCGGAG	1525
Db		1260	GGCCTGACAGGACCACTGGCTCAGGTGCCCTGTCTCTGGGCGGAGCACCTGCGGAG	1319
OY		1526	CTGCCCCTCTTGGTGTGTATGAAGACCAAGGCCGCGGACCGGTTCAGAGACATGAGAC	1585
Db		1320	CTGCCCCTCTTGGTGTGTATGAAGACCAAGGCCGCGGACCGGTTCAGAGACATGAGAC	1379
OY		1586	CACCACACCATGAAACATCACCACTATTACCACTTCTACAGACATAAGAGCCCTCC	1645
Db		1380	CACCACACCATGAAACATCACCACTATTACCACTTCTACAGACATAAGAGCCCTCC	1439
OY		1646	CCAGGCGCCCAACCCCTGCCATATGAAGAGCCCAACCCCGACACCAAGGCAATTATTT	1705
Db		1440	CCAGGCGCCCAACCCCTGCCATATGAAGAGCCCAACCCCGACACCAAGGCAATTATTT	1499
OY		1706	CTATTTAATTATGTTATATATATATATATGTTAATTATTAATTGTTACTCCACTAAT	1765
Db		1500	CTATTTAATTATGTTATATATATATATATGTTAATTATTAATTGTTACTCCACTAAT	1559
OY		1766	ATTTAGCTAGCCTACATGTAGAAAGATCTATGAAAACACAGAACTTAACTTTATTTAT	1825
Db		1560	ATTTAGCTAGCCTACATGTAGAAAGATCTATGAAAACACAGAACTTAACTTTATTTAT	1619
OY		1826	GTT 1828	
Db		1620	GTT 1622	
 RESULT 10 US-10-087-192-178 Sequence 178, Application US/10087192 Publication NO. US20020182586A1 GENERAL INFORMATION: APPLICANT: Morris, David W. APPLICANT: Engelhard, Eric K. TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR FILE REFERENCE: 529452000122 CURRENT APPLICATION NUMBER: US/10/087,192 CURRENT FILING DATE: 2002-03-01 PRIOR APPLICATION NUMBER: US 09/747,377 PRIOR FILING DATE: 2000-12-22 PRIOR APPLICATION NUMBER: US 09/798,586 PRIOR FILING DATE: 2001-03-02 NUMBER OF SEQ ID NOS: 2059 SOFTWARE: FastSeq for Windows Version 4.0 SEQ ID NO 178 LENGTH: 106315 TYPE: DNA ORGANISM: Homo sapiens FEATURE: NAME/KEY: misc_feature LOCATION: (1)...(106315) OTHER INFORMATION: n = A,T,C or G US-10-087-192-178				

Query Match	42.1%	Score 783;	DB 13;	Length 106315;
Best Local Similarity	100.0%	Pred. No. 4.1e-176;		
Matches 783; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

OY	1046	GAGCTCCCTTCGCGTGGCCCAAGAAGTCAGAATGCCCCGCCGCACCTCACAATCCACTCGA	1105
Dd	95728	GAGCTCCCTTCGCGTGGCCCAAGAAGTCAGAATGCCCCGCCGCACCTCACAATCCACTCGA	9578
OY	1106	TCTCGCTCCCATGAGCGCGAAGCCATCCACATCCCAACCGAAAGCCCACAAGCGTGGAAC	1165
Dd	95788	TCTCGCTCCCATGAGCGCGAAGCCATCCACATCCCAACCGAAAGCCCACAAGCGTGGAAC	9584
OY	1166	CCGGCCCTTCCTTCCAATTCTTTGACACCCCAATGGCCAAAGGTCTCAGAGCTCCAGCAACGG	1225
Dd	95848	CCGGCCCTTCCTTCCAATTCTTTGACACCCCAATGGCCAAAGGTCTCAGAGCTCCAGCAACGG	9590
OY	1226	CTCCGGGGACAACCAGAGCGGAGAACACATTTGTGAGTCTCCCCCAAGGCCCAAGGCAAG	1285
Dd	95908	CTCCGGGGACAACCAGAGCGGAGAACACATTTGTGAGTCTCCCCCAAGGCCCAAGGCAAG	9596
OY	1286	AGTGTGGGTGTGGGCCACAGTGCCCAAGGGGCAGAAACAAGCCCTCTTGAGAACCCGCC	1345
Dd	95968	AGTGTGGGTGTGGGCCACAGTGCCCAAGGGGCAGAAACAAGCCCTCTTGAGAACCCGCC	9602
OY	1346	ATCCCTGCGGATGTCCCTCTCGGCCACATGAGCTGSCAGGCCCGGCCCTCTCCCTCCCTTA	1405
Dd	96028	ATCCCTGCGGATGTCCCTCTCGGCCACATGAGCTGSCAGGCCCGGCCCTCTCCCTCCCTTA	9608
OY	1406	GCCCCCTCGGGGCACAAGAAAGCACAAAGCACGAGCCAGAGAGCCAGCAGAGCGCTGCGG	1465
Dd	96088	GCCCCCTCGGGGCACAAGAAAGCACAAAGCACGAGCCAGAGAGCCAGCAGAGCGCTGCGG	9614
OY	1466	GAGCTGCAAGGACCACTGGGCTCAGGTGGCCCTGTCTCGGGGGGGAGACCTGCGGGAG	1525
Dd	96148	GAGCTGCAAGGACCACTGGGCTCAGGTGGCCCTGTCTCGGGGGGGAGACCTGCGGGAG	9620
OY	1526	CTGCCCGCCTTGGTGGTGTATGAGAGCGAGCGGAGCGGCTCAAGACATAGACAC	1585
Dd	96208	CTGCCCGCCTTGGTGGTGTATGAGAGCGAGCGGAGCGGCTCAAGACATAGACAC	9626
OY	1586	CACCAACCAATGAAATCAACCACTTACCACTTCTTACAGACATAGAGCCCTCC	1645
Dd	96268	CACCAACCAATGAAATCAACCACTTACCACTTCTTACAGACATAGAGCCCTCC	9632
OY	1646	CCAGGGCCCCACCTGCGCATATGAAAGACCCCAACCCCGGACACACAAGGACATTAATAT	1705
Dd	96328	CCAGGGCCCCACCTGCGCATATGAAAGACCCCAACCCCGGACACACAAGGACATTAATAT	9638
OY	1706	CTATTAAATTAATGTTATATGATGATTAATTTATTAATAATTAATGTTACTCCACTAAT	1765
Dd	96388	CTATTAAATTAATGTTATATGATGATTAATTTATTAATAATTAATGTTACTCCACTAAT	9644
OY	1766	ATTTAGCTAGCCTACATGTAGAAAGATCTATGAAAACAGAACTAAACTTTTATTTATAT	1825
Dd	96448	ATTTAGCTAGCCTACATGTAGAAAGATCTATGAAAACAGAACTAAACTTTTATTTATAT	9650
OY	1826	GTT	
Dd	96508	GTT 96510	

```

RESULT 11
US-09-993-966-21
; Sequence 21, Application US/09993966
; Publication No. US20030186232A1
; GENERAL INFORMATION:
; APPLICANT: ROHAN, MICHAEL
; TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF MKD PROTEIN
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF
; FILE REFERENCE: 014024/028073
; CURRENT APPLICATION NUMBER: US/09/993,966
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 60/252,884
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/291,109
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/325,571

```



```

; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8686
; LENGTH: 427
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(427)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-8686

```

```

Query Match      22.5%; Score 418.2; DB 10; Length 427;
Best Local Similarity 99.3%; Pred. No. 7,7e-90;
Matches 420; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1076 CTGCCCCCGGCACTTCCATTCCTGATCTGCTTCCATGAGCCGGAAGCCATCCAC 1135
DB 4 CTGCCCCCGGCACTTCCATTCCTGATCTGCTTCCATGAGCCGGAAGCCATCCAC 63
QY 1136 ATCCCAACGGAAGAGCCGGAAGGCTGGAACCCGAGCTCTTCCACTTCTTGAACCCCA 1195
DB 64 ATCCCAACGGAAGAGCCGGAAGGCTGGAACCCGAGCTCTTCCACTTCTTGAACCCCA 123
QY 1196 ATGCGCAAGGCTTCAAGCTTCAAGCAAGGCTCCGAGGCAACCAAGAGGAGCAAC 1255
DB 124 ATGCGCAAGGCTTCAAGCTTCAAGCAAGGCTCCGAGGCAACCAAGAGGAGCAAC 183
QY 1256 TTTGTAGGTCTCCCAAGGCGGCAAGAGTGTGGTGTGGGCACTGTGGCAAGGG 1315
DB 184 TTTGTAGGTCTCCCAAGGCGGCAAGAGTGTGGTGTGGGCACTGTGGCAAGGG 243
QY 1316 GCAAGAAACAGAGCCCTCTGGAACCGGCACTCTGAGGCTCCGAGGAGGCTCCGAGGAG 1375
DB 244 GCAAGAAACAGAGCCCTCTGGAACCGGCACTCTGAGGCTCCGAGGAGGCTCCGAGGAG 303
QY 1376 GCTGCAAGCCGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1435
DB 304 GCTGCAAGCCGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 363
QY 1436 CGAGCCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1495
DB 364 CGAGCCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 423
QY 1496 CCT 1498
DB 424 CCT 426

```

```

RESULT 14
US-09-764-868-152
; Sequence 152, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 152

```

```

; LENGTH: 1966
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-152

```

```

Query Match      14.7%; Score 272.4; DB 9; Length 1966;
Best Local Similarity 56.5%; Pred. No. 1.2e-54;
Matches 667; Conservative 0; Mismatches 471; Indels 42; Gaps 7;

```

```

QY 497 GAGAAAGCTGACGGGCTGGGACCGGAGATGAGAAAGATGAGAGATGAGGAAACC 556
DB 516 GAGAACCCGGGACAACTCTCAGCGAGATGACGAGAGAGGAGGAGGAGGAGGAGGAGG 575
QY 557 TGCCCAAGGCTC---CAAGAGGAGCTGAATTTGAAGAGCTCCAGTCCAGCTGTCCATG 613
DB 576 CCGGAGAGACCGGAGCGGAGCGGCTTCAATTGACGACTTCCAGTCCAGTCCAGTCCG 635
QY 614 GAGAGGAGACCGGAGCGGAGGAGTGAACCTTCACTTGTATGACTTTGACAAACGCGAG 673
DB 636 GAGAGGAGACCGGAGCGGAGGAGTGAACCTTCACTTGTATGACTTTGACAACTGCGGAG 695
QY 674 GTCAACCGAGAGGACATCAACAGCTTGTGACACACATCTATGAGGTGTGACTCTCT 733
DB 696 GTCAACCGAGAGGACATCTGACGCTCATGACACATCTATGAGGTGTGACTCTCTCT 755
QY 734 GTCAACCTCTCCCAACATCTCAGCAAGATGCTGCGGTAAAGCTCACTGTGCGCCGAT 793
DB 756 GTCAACCTCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 815
QY 794 GGCAGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 853
DB 816 CCTTCCAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 875
QY 854 CGAGCAGAGACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 913
DB 876 AGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 935
QY 914 CTGAGGTTCCAGGAGTGAACGCGGCTGTGAGCACTGTGCTGCTGCTGCTGCTGCTG 973
DB 936 GTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 995
QY 974 GAGAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1033
DB 996 GAGAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1055
QY 1034 CAATTTGGGCTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1093
DB 1056 AGATTTGGGCTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1115
QY 1094 AATCCCA---CTGATCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1150
DB 1116 CACTTCAAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1175
QY 1151 CCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1210
DB 1176 TCACAGGTGTGTGAGAAACATCTGTGCAAGCTTCGAA---GCTGTGCTCCGAGGCTCTG 1232
QY 1211 GAGCTCAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1270
DB 1233 GACACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1283
QY 1271 AAGGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1330
DB 1284 AAGGGCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1343
QY 1331 CTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1390
DB 1344 TACTACTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1397
QY 1391 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1450
DB 1398 CCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1457

```



```
OY 1451 CAGCAGGGCTGCGGGGCTGTGACAGCACACTGCGCTCAGATGGCCCTGTCTTGGGGCGG 1510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1458 CACTCGCCACTCAAGAGGCCCAACAGCTCAGCCTTGCACAG-----TGGAGCAC 1505
OY 1511 GAGCAGCTGCGGGAGCTGCGCCGCTTGGTGTATGAGACCAAGCCGGGCGAGCCGTC 1570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1506 GAGGTGGTGGGGCTGCGCGCCCAAGCCAGAGAGGGCTTACGCGTGCAGTATC 1565
OY 1571 CAGAGCATAGACCAACCAACCAATGAACATACCAACATT-----ACCACTTTC 1624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1566 CAGGGACAGACCAACCAACCAAGACCAACCAACCAACCAACCAACCAACCACTTC 1625
OY 1625 TACCAGCATAGAGCCCTCCCGAGGGCCCAACCTGCCA 1664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1626 CACCCGCTAGCGCCCACTGCAAGCACACCTCGCTCCCA 1665
```

```
RESULT 15
US-09-993-966-22
/ Sequence 22, Application US/09993966
/ Publication No. US20030186232A1
/ GENERAL INFORMATION:
/ APPLICANT: ROHAN, MICHAEL
/ TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NKD PROTEIN,
/ TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF
/ FILE REFERENCE: 014024/0280733
/ CURRENT APPLICATION NUMBER: US/09/993,966
/ CURRENT FILING DATE: 2001-11-27
/ PRIOR APPLICATION NUMBER: 60/252,884
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/291,109
/ PRIOR FILING DATE: 2001-05-16
/ PRIOR APPLICATION NUMBER: 60/325,571
/ PRIOR FILING DATE: 2001-10-01
/ NUMBER OF SEQ ID NOS: 26
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 22
/ LENGTH: 215
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-993-966-22
```

```
Query Match 11.6%; Score 215; DB 10; Length 215;
Best Local Similarity 100.0%; Pred. No. 2.5e-41;
Matches 215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1637 AGCCCTCCCGAGGCCCCCACTGCCATATGAAGACCCCGACCAACCAAGGC 1696
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 AGCCCTCCCGAGGCCCCCACTGCCATATGAAGACCCCGACCAACCAAGGC 60
OY 1697 ATTATTAATTCATTAATTATTTGTTATATGATGATTAATTTGTTAATTAATTAATTTGTTAC 1756
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 ATTATTAATTCATTAATTATTTGTTATATGATGATTAATTTGTTAATTAATTAATTTGTTAC 120
OY 1757 TCCACTAATATTTTACTAGCTTACATGTAGAGATCTATGGAACAACAGAACTAACTTT 1816
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 TCCACTAATATTTTACTAGCTTACATGTAGAGATCTATGGAACAACAGAACTAACTTT 180
OY 1817 TATTATATGTGTAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1851
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 TATTATATGTGTAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 215
```

Search completed: December 29, 2004, 18:07:29
Job time : 5339 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 29, 2004, 13:06:18 ; Search time 8793 Seconds
(without alignments)
7704.022 Million cell updates/sec

Title: US-09-993-966-5

Perfect score: 1859

Sequence: 1 gaattcgccctctactacg.....aaaaaaaaaagcgcgccg 1859

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Database : EST.*
Listing first 45 summaries

1: gb_esc1:*
2: gb_esc2:*
3: gb_esc3:*
4: gb_esc4:*
5: gb_esc5:*
6: gb_esc6:*
7: gb_esc7:*
8: gb_esc8:*
9: gb_esc9:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1753.2	94.3	2142	3	AF289584 Homo sapi
2	1242.8	66.9	2289	3	AK082367 Mus muscu
3	1197.2	64.4	2471	3	AK076164 Mus muscu
4	1027.6	55.3	1466	3	AK013241 Mus muscu
5	885	47.6	921	5	AY412098 Homo sapi
6	825.6	44.5	921	5	BQ645656 AGENCOURT
7	825.6	44.4	931	5	BQ653673 AGENCOURT
8	810.6	43.6	990	5	BQ064678 AGENCOURT
9	793	42.7	928	5	BQ064956 AGENCOURT
10	789.4	42.5	895	5	BQ644360 AGENCOURT
11	757.6	40.8	953	5	BQ846054 AGENCOURT
12	748.8	40.3	985	5	AY412099 Pan trogl
13	712.6	38.3	966	5	BQ649813 AGENCOURT
14	711.4	38.0	996	5	BQ652087 AGENCOURT
15	706	38.0	1008	5	BQ646371 AGENCOURT
16	700.4	37.7	1027	4	BG104777 602311726
17	689.6	37.1	1059	5	AY412100 Mus muscu
18	682.6	36.7	1059	5	BQ070932 AGENCOURT
19	661.6	35.6	825	6	CB961961 AGENCOURT
20	622	33.5	622	6	CB215756 NISC np09
21	611.8	32.9	899	5	BQ645507 AGENCOURT
22	597.8	32.2	826	6	CB961668 AGENCOURT
23	593.6	31.9	913	5	BQ895164 AGENCOURT
24	593	31.9	634	7	CN369119 170005999

25	592.2	31.9	777	4	BG542261 602571809
26	584.2	31.4	627	4	B1767278 603057995
27	583.2	31.4	915	6	CB182455 AGENCOURT
28	561.4	30.2	564	4	BM711145 UI-E-DX1-AG112447 Pan trogl
29	549.2	29.5	654	9	AG112447 Pan trogl
30	532	28.6	677	7	CF731367 UI-M-G20-CB05689 NISC j119
31	530	28.5	674	6	CB05689 NISC j119
32	509.2	27.4	946	6	B1712252 UI-M-FY0-CA32830 B1712252
33	508.2	27.3	724	6	CA32830 UI-M-FY0-CK795599 AGENCOURT
34	503.8	27.1	839	7	BG820139 602782356
35	477.4	25.7	751	4	BG820139 M83-FN020
36	467.4	25.1	481	4	B1047069 M83-FN020
37	466.2	25.1	678	2	BF537335 602053266
38	459.4	24.7	653	2	BF540602 602054683
39	458	24.6	670	5	BQ554861 H4030603-BB650914 BB650914
40	456.6	24.6	670	5	CO811260 AGENCOURT
41	455.4	24.5	693	7	BM935129 UI-M-BH3-BU856404 AGENCOURT
42	452.6	24.3	693	5	BU856404 AGENCOURT
43	451.2	24.3	931	5	BU856404 AGENCOURT
44	431	23.2	817	7	CK790579 AGENCOURT
45	425.8	22.9	429	7	CN369118 170005322

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AF289584	AF289584	Homo sapiens clone pp7246 unknown mRNA.	2142 bp	mrna	linear	HTC 01-JAN-2002	
AF289584	AF289584	GI:18027371					
HTC							
Homo sapiens							
Homo sapiens (human)							
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.							
1 (bases 1 to 2142)							
Zhang, P.P., Zhou, X.M., Jiang, H.Q., Huang, Y., Qin, W.X., Zhao, X.T., Wan, D.F. and Gu, J.R.							
Novel human cDNA clones with function of inhibiting cancer cell growth							
Unpublished							
2 (bases 1 to 2142)							
Zhang, P.P., Zhou, X.M., Jiang, H.Q., Huang, Y., Qin, W.X., Zhao, X.T., Wan, D.F. and Gu, J.R.							
Direct Submission							
Submitted (17-JUN-2000) National Laboratory For Oncogenes & Related Genes, Shanghai Cancer Institute, 25/Jun 2200 Xie-Tu Road, Shanghai 200032, P. R. China							
Location/Qualifiers							
1. 2142							
/organism="Homo sapiens"							
/mol_type="mrna"							
/db_xref="taxon:9606"							
/clone="pp7246"							
277. 1689							
/codon_start=1							
/product="unknown"							
/protein_id="AA155768.1"							
/db_xref="GI:18027372"							
/translation="MGKLSKPDVCKRRESPEGDSFAVSAMARGIIEWIGRCPGGVSGRQLRAGTIGRSTREIVGLVRLTLEBEDDRLLEALPPEKTDIGSGDEKMERVSEPCSKQLKEELIQCDVMEDESRQEWTFLLYFDNDNGASAPRAETKETEHTIYEVDSVNSHSPTSKMLRVKLTVAVDGOSKRSVAVNODLDSAPRAETKETE DLSWEKKORAPLRDGRRLRSGCYHHCQVDENTERRHHVLDIAIEVYTSFGGSGPSVAQSELRPTSNTRSRSHPEPAIHIPHRKPGCVDPASRFRFDTPIAKSELDQRLGTDGSGHFAVRSVAQSGVGVGVHVAAGANKKPLGPAIVAVSPSHLASPALPLSLAPLGKKGKIRAKESQOQCGRLQAPLASGPGVLRGRRLRLPALVVESSQAGDPVQ RHHHHHHHHHHHHFYQT"							
ORIGIN							
Query Match	94.3%	Score 1753.2	DB 3	Length 2142			

Best Local Similarity 99.5%; Pred. No. 0;
Matches 1758; Conservative 0; Mismatches 8; Indels 0; Gaps 0.

OY	63	GGGAGTCGGGGCCGGCGGACGCGCGCAGAGACCGGCTCCCGGCGCGCTCGGCGCTCCGCT	122
Db	116	GTGAGTCGGGGCCGGCGAGACGAGCGCGCAGAGACCGGCTCCCGGCGCGCTCGGCGCTCGGCT	175
OY	123	CGGCTCGGGGGCTGCTTCGGGAGGAGAGAGACCAAGGAGGCGCGCAGGCTCCGCGGCGCG	182
Db	183	GCGCATGGCTTAAGGAGCGCTCCCGGCGCGCGCAGGCTCCGCGGCGCGCTCGGCGCTCGA	242
OY	236	GCGCATGGCTTAAGGAGCGCTCCCGGCGCGCGCAGGCTCCGCGGCGCGCTCGGCGCTCGA	295
OY	243	AGCGGCGCGCGGTGCGCAAGCGCAGAGGAGACCCGGAAAGTGAACAGCTTCGCGCGTGAAGC	302
Db	296	AGCGCGTGAAGTGTGCAAGCGCAGAGGAGACCCGGAAAGTGAACAGCTTCGCGCGTGAAGC	355
OY	303	CTGCGTGGGCTTCGGAAGGCGATCCGAGAGTGAATCGGAGACAGCGCTGCGCGGCGGTG	362
Db	356	CTGCGTGGGCTTCGGAAGGCGATCCGAGAGTGAATCGGAGACAGCGCTGCGCGGCGGTG	415
OY	363	TCTCGGAGACCCCGACAGCTGCGGTTGGCGGGGCAACATGAGGCGGAAGCACCCGGGAGCTCG	422
Db	416	TCTCGGAGACCCCGACAGCTGCGGTTGGCGGGGCAACATGAGGCGGAAGCACCCGGGAGCTCG	475
OY	423	TGGGCGACGTGTGAGAGACAGCGCTCAGCGAGAGAGAGAGAGACGACTTTGCGCTGGAAG	482
Db	476	TGGGCGACGTGTGAGAGACAGCGCTCAGCGAGAGAGAGAGAGACGACTTTGCGCTGGAAG	535
OY	483	TGGGCGCTGCTCTCGAGAGAGACTGAACGGGCTGAGCGAGATGAGAGAGAGATGAGAG	542
Db	536	TGGGCGCTGCTCTCGAGAGAGACTGAACGGGCTGAGCGAGATGAGAGAGAGATGAGAG	595
OY	543	GAGTGAAGGAAACCTGCGCCAGGCTTCAGAGAGACAGCTGAAGTTTGAAGAGCTCCAGTGC	602
Db	596	GAGTGAAGGAAACCTGCGCCAGGCTTCAGAGAGACAGCTGAAGTTTGAAGAGCTCCAGTGC	655
OY	603	AACGTTCATGAGAGAGAGACAGCCGGGAGAGTGAACCTTCAACCTGTATGACTTTGACA	662
Db	656	AACGTTCATGAGAGAGAGACAGCCGGGAGAGTGAACCTTCAACCTGTATGACTTTGACA	715
OY	663	ACAACGGCAAGGTCACCCGAGAGAGACATCACAGCTTGCTGACACACATCTATGAGGTG	722
Db	716	ACAACGGCAAGGTCACCCGAGAGAGACATCACAGCTTGCTGACACACATCTATGAGGTG	775
OY	723	TGGAATCCTCTGTCAACCACTCCCAATCAAGCAATGCTGTGGGTAAAGCTCACCG	782
Db	776	TGGAATCCTCTGTCAACCACTCCCAATCAAGCAATGCTGTGGGTAAAGCTCACCG	835
OY	783	TGGCCCCCGATGAGCAGCCAGAGCAAGAGAGAGTCTTGTCAATCAGGCTGACCTGCGAGA	842
Db	836	TGGCCCCCGATGAGCAGCCAGAGCAAGAGAGAGTCTTGTCAATCAGGCTGACCTGCGAGA	895
OY	843	GCGCAAGGCTCCGAGCAGAGACCAAGCCCACTGAGAGACTTCGCGAGCTGGAGAGAAAGC	902
Db	896	GCGCAAGGCTCCGAGCAGAGACCAAGCCCACTGAGAGACTTCGCGAGCTGGAGAGAAAGC	955
OY	903	AGCGAGCCCGGCTCAGTTCCAGGGGTGACGCGGCTGAGACAGCTGAGCTGATACACG	962
Db	956	AGCGAGCCCGGCTCAGTTCCAGGGGTGACGCGGCTGAGAGAGCTTGGCTGATACACG	1015
OY	963	ATTGCGTATGAGAAACATCGAGAGAGAAACCACTAATTAGATTCGCGCGGATAGAA	1022
Db	1016	ATTGCGTATGAGAAACATCGAGAGAGAAACCACTAATTAGATTCGCGCGGATAGAA	1075
OY	1023	ACTACAGTCCCAATTTGGGCTGGCTCCCTTCGTTGGCCCAAGAGTCAAGACTGCGCC	1082
Db	1076	ACTACAGTCCCAATTTGGGCTGGCTCCCTTCGTTGGCCCAAGAGTCAAGACTGCGCC	1139
OY	1083	CCCGAGCTTCATTCAGATTCGATTCGCTCCCAAGACCGGAAAGCAATCAGATCCAC	1142

Db	1136	CCCGCACTCCAACTCCAACTCGATCTCGCTCCATGAGCCGGAAGCCATCCATCTCCAC	1195
Qy	1143	ACCCAAAGCCCCCAAGAGCGTGGACCCCGGCTCTCTTCCACTTCTCTTGACACCCCAATGCGCA	1202
Db	1196	ACCGAAGAGCCCAAGAGGCTGGACCCCGGCTCTTCCACTTCTCTTGACACCCCAATGCGCA	1255
Qy	1203	AGGTTCTGAGACTCCAGCAACGCTCCGGGGCAACCCAGAGCGGAGCAAGCATTTTGTGA	1262
Db	1256	AGGTTCTGAGACTCCAGCAACGCTCCGGGGCAACCCAGAGCGGAGCAAGCATTTTGTGA	1315
Qy	1263	GGTCCCCCAAGGCCCAAGGCAAGGTGCGGGTGAGGCAAGTGAGCCAGAGGCAAGAA	1322
Db	1316	GGTCCCCCAAGGCCCAAGGCAAGGTGCGGGTGAGGCAAGTGAGCCAGAGGCAAGAA	1375
Qy	1323	ACAAGCCCCCTCTGGAGACCGGCATCTCCCTGGGTGTCCCTCTCGGCCACTGTGGCTGCA	1382
Db	1376	ACAAGCCCCCTCTGGAGACCGGCATCTCCCTGGGTGTCCCTCTCGGCCACTGTGGCTGCA	1435
Qy	1383	GCCCCGCCCTCTCTCCCTCCCTAGCCCCCTCGGGCAAGAGAGCAAGAGCAAGAGCA	1442
Db	1436	GCCCCGCCCTCTCTCCCTCCCTAGCCCCCTCGGGCAAGAGAGCAAGAGCAAGAGCA	1495
Qy	1443	AGGAGAGCCAGCAGAGGCTGCCCGGGGCGCTGAGAGACACATGAGCCTCAGGTGAGCCCTGTC	1502
Db	1496	AGGAGAGCCAGCAGAGGCTGCCCGGGGCGCTGAGAGACACATGAGCCTCAGGTGAGCCCTGTC	1555
Qy	1503	TGGGCGCGGAGCAGCTGCGGGAGCTGCGCGCTTGTGTGTGTATGAGAGCCAGGCGGGC	1562
Db	1556	TGGGCGCGGAGCAGCTGCGGGAGCTGCGCGCTTGTGTGTGTATGAGAGCCAGGCGGGC	1615
Qy	1563	AGCCGGTCCAGAGCATGAGGACCAACACCAACATGAAATCACACACTTATCCACACT	1622
Db	1616	AGCCGGTCCAGAGCATGAGGACCAACACCAACATGAAATCACACACTTATCCACACT	1675
Qy	1623	TCTACAGACATTAAGCCCCCTCCCAAGGGCCCCACCTGCGATGTGAAGAGCCCAACCC	1682
Db	1676	TCTACAGACATTAAGCCCCCTCCCAAGGGCCCCACCTGCGATGTGAAGAGCCCAACCC	1735
Qy	1683	CGACACCAAGAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1742
Db	1736	CGACACCAAGAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1795
Qy	1743	ATAATTATGTACTCCATATAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1802
Db	1796	ATAATTATGTACTCCATATAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1855
Qy	1803	CAGAACTTAACCTTTATTTATTAATGTT	1828
Db	1856	CAGAACTTAACCTTTATTTATTAATGTT	1881

RESULT 2	
AK082367	
LOCUS	2289 bp mRNA linear HTC 03-APR-2004
DEFINITION	Mus musculus 0 day neonate cerebellum cDNA; RIKEN full-length enriched library, clone:C30040009 product:naked cuticle 1 homolog (Drosophila), full insert sequence.
ACCESSION	AK082367
VERSION	AK082367.1 GI:26100612
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	
AUTHORS	1 Carninci, P. and Hayashizaki, Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Mech. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	2
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, H., Konno, H., Okaaki, Y., Muramatsu, M. and Hayashizaki, Y.

Db 953 CTACAGCTCAAGTTTGAGCCGGGATCCCTTCGCTGGTGGCCAGAAAGTCAGAGCTGCCCC 1012
 QY 1084 CCGGACCTCCAAATCCCATCTGATTCGCTCCCATGAGCCGGAAGCCATCCACATCCACA 1143
 Db 1013 TCGAATTTCCAAACCCCACTGCTCTGCTCCCAAGACCAGAAAGCTCCCAATCCACA 1072
 QY 1144 CCGAAAGCCCAAGGCGTGAACCCGCGCTCTTCACTTCTTGAACACCCCAATCGCAA 1203
 Db 1073 CCGGAGGCCCCCAAGTGTGAGACCAAGGCTCTTCCACTCTCTTGAACACCCCAATTTGCCAA 1132
 QY 1204 GGTCTCAGAGCTCCAGAACGCTCCGCGGCAACCAAGAGGAGCAAGCACTTTGTGAG 1263
 Db 1133 GGCAATCAGAGCTCCAGAACGCGCTCCGGGGCACTCAGATGGAGGACAAGCACTTTGTGAG 1192
 QY 1264 GTCCCCCAAGGCCCAAGAGTGTGGGTGTGGGCACTGTGGCCAGAGGGGCAAGAAA 1323
 Db 1193 GTCCCCCAAGGCCCAAGAGTGTGGGTGTGGGCACTGTGGCCAGAGAGTGTGAAAG 1252
 QY 1324 CAAGCCCTCTGGGAGCCGCGCATCCCTGGGTGTCCCTCCGCGCCACCTGGCTGGCAG 1383
 Db 1253 CAAGCTCCACTGGTACCACCACTACTGTCTCCCTCTTGCCTTGTGGCCACAG 1312
 QY 1384 CCGGACCTCTCTCCCTCCCTAGCCCTCCGCGCAAGAGCAAGCAAGCAAGCAAGCAAGCA 1443
 Db 1313 CCGAGCCCTTCTCCCTCCCTAGCCCTCCGCGCAAGAGCAAGCAAGCAAGCAAGCAAGCA 1372
 QY 1444 GAGAGCCAGCAGGCTGTGGGGGCTGTGAGGACCACTGGGCTCAAGTGG---CCCTGT 1500
 Db 1373 GAGAGCCAGCAGGCTGTGGGGGCTGTGAGGACCACTGGGCTCAAGGCTTCACCGT 1432
 QY 1501 CTTGGGCGGGAGCAGCTGGCGGAGTGTCCGCTGTGGTGTATGAGAGCCAGCGG 1560
 Db 1433 CATGGGCGGGAGGAGGTGAGGAGTGTCTGCTGCTGTGTGTATGAGAGCCAGGCTGG 1492
 QY 1561 GCAAGCCGCTCCAGAGCATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1620
 Db 1493 GCAAGCCGCTCCAGAGCATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1552
 QY 1621 CTTCTACAGCATGAGAGGCTGTGGGGGCTGTGAGGACCACTGGGCTCAAGTGG---CCCTGT 1680
 Db 1553 CTTCTACAGCATGAGAGGCTGTGGGGGCTGTGAGGACCACTGGGCTCAAGTGG---CCCTGT 1601
 QY 1681 CCGGACCCAGCAGGCTGTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1740
 Db 1602 C---ACACCTTAAGGCTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1658
 QY 1741 TAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1800
 Db 1659 TAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1718
 QY 1801 CACAGAACTAACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1828
 Db 1719 CACAGAACTAACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1746

RESULT 3
 LOCUS AK076164 2471 bp mRNA linear HTC 03-APR-2004
 DEFINITION Mus musculus 14, 17 days embryo head cDNA, RIKEN full-length
 enriched library, clone:330401D10 product:naked cuticle 1 homolog
 (Drosophila), full insert sequence.
 ACCESSION AK076164
 VERSION AK076164.1 GI:26096648
 SOURCE HTC; CAP trapper.
 KEYWORDS Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253

PUBMED 10349636
 REFERENCE 2
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE 3
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Komno, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujitake, S., Inoue, K., Togawa, Y., Izawa, Y., Ohara, E., Matsubara, S., Kawai, J.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system-384-format
 JOURNAL sequencing pipeline with 384 multichipillary sequencer
 MEDLINE Genome Res. 10 (11), 1757-1771 (2000)
 PUBMED 20530913
 REFERENCE 4
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 REFERENCE 5
 AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 JOURNAL of 60,770 full-length cDNAs
 REFERENCE Nature 420, 563-573 (2002)
 TITLE 6 (bases 1 to 2471)
 REFERENCE Aachi, U., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Onno, M., Ohnato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takekida, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp,
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 COMMENT cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.jp/
 URL: http://fantom.gsc.riken.jp/
 URL: http://location.qualifiers

FEATURES
 source
 1. 2471
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM:DB:3230401D10"
 /db_xref="taxon:10090"
 /clone="3230401D10"
 /tissue_type="head"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="14, 17 days embryo"

REFERENCE	AUTHORS	Carminci, P. and Hayashizaki, Y.
TITLE	TITLE	High-efficiency full-length cDNA cloning
JOURNAL	JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	MEDLINE	99279253
PUBMED	PUBMED	10349636
REFERENCE	AUTHORS	2
TITLE	TITLE	Carminci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komoto, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
JOURNAL	JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
MEDLINE	MEDLINE	Genome Res. 10 (10), 1617-1630 (2000)
PUBMED	PUBMED	20499374
REFERENCE	AUTHORS	3
TITLE	TITLE	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carminci, P., Komoto, H., Akiyama, U., Nishi, K., Kitunai, T., Tashiro, H., Itoh, A., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, K., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
TITLE	TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipicillary sequencer
JOURNAL	JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	MEDLINE	20530913
PUBMED	PUBMED	11076861
REFERENCE	AUTHORS	4
TITLE	TITLE	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
JOURNAL	JOURNAL	Functional annotation of a full-length mouse cDNA collection
MEDLINE	MEDLINE	Nature 409, 685-690 (2001)
PUBMED	PUBMED	5
REFERENCE	AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	TITLE	Analysis of the mouse transcriptome based on functional annotation
JOURNAL	JOURNAL	of 60,770 full-length cDNAs
MEDLINE	MEDLINE	Nature 420, 563-573 (2002)
PUBMED	PUBMED	6 (bases 1 to 1466)
REFERENCE	AUTHORS	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carminci, P., Fukuda, S., Fukunishi, K., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hizumoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasehwa, T., Kato, C., Kawai, U., Kojima, Y., Komoto, H., Kouda, M., Koyu, S., Kuzihara, C., Matsumura, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ono, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Teijima, Y., Toyai, T., Yamamura, T., Yaunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.
TITLE	TITLE	Direct Submission
JOURNAL	JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki. The Institute of Physiol and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	COMMENT	please visit our web site (http://genome.gsc.riken.jp/) for further details.
		cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in RIKEN Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGATTCAGTTAAATTAATGCCCCCCCC 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 7.5 and subtraction to Rot = 37.5. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCAGTTAAATTAATGCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI, 3' end: SstI.

FEATURES					
Source					
Host: SOLR.					
Location/Qualifiers					
1..1466					
/organism="Mus musculus"					
/mol_type="mrna"					
/strain="C57BL/6J"					
/db_xref="PANTOM.DB:2810434U10"					
/clone="2810434U10"					
/issue_type="whole body"					
/clone_lib="RIKEN full-length enriched mouse cDNA library"					
/dev_Stage="10, 11 days embryo"					
1..1466					
/note="naked cuticle 1 homolog (Drosophila)					
(MG MG1:2135954, GB NM_027280, evidence: BLASTN, 98%, match=1658)"					
ORIGIN					
Query Match 55.3%, Score 1027.6; DB 3, Length 1466,					
Best Local Similarity 84.7%; Pred. No. 7.8e-210;					
Matches 1194, Conservative 0, Mismatches 199, Indels 17, Gaps 3					
OY	415	GGAGCTGTTGGCGACGTTGTGAAGACAGCTCAGCAGGAGAAGAGACACTTTCG	474		
Db	71	GGAACTCGTGGGTACACTTCTAAGAGGCTCTCGGTAGAGAGACAGACGACTTCCC	130		
OY	475	GCTGAAATGCGCCCTGCCTCTTGAGAGACTGACGGGCTGGGCAACGAGATGAGAGAA	534		
Db	131	CCTAGAAATGCCCCCTGCCGCTGAGAAATTGACAGCCTTAGGTAGATGAGAGAG	190		
OY	535	GATGAGAGAGTGAACCGAACCTTCGCCAGGCTCCAAGAGCAGCTGAAGTTTGAAGACT	594		
Db	191	AATGGAGAGACTGAGCGCACTTGCCACAGGCTCCAGAGAGAGCTCAAGTTTGAAGACT	250		
OY	595	CCAGTGGCAGCTGTCATGAGAGAGACAGCGGAGAGAGTGAACCTCACCTGTATGA	654		
Db	251	GCAGTGATATCTCTGTGAGAGAGACAGCCGGCAGAGATGACTTTCATCTATATGA	310		
OY	655	CTTTGACAACAACGGCAAAGTCAACCAGAGAGATCAACGCTTGCTGCAACCATCTA	714		
Db	311	CTTGAGACAACATGGCAAAAGTGAACCCGTAAGAGACATTAACGACTTGCTGACATACATCTA	370		
OY	715	TGAGGTGTGACTCTCTGTCAACCACTCCCCAATCCAGCAAGATGCTGCGGGTAAA	774		
Db	371	TGAAGTGTGACTCTCTGTGAACCAATCCCCACATCAAGCAAGACATGCGGGTAAA	430		
OY	775	GCTCAACSTGGCCCCCATGGCAGCCAGAGCAAGAGAGAGCGCTCTGTCAATCAGACTGA	834		
Db	431	GCTCAACATGGCTCTTGAACGGAGCCAGAGTAAGAGAGCGCTCTTTCAACATAACGA	490		
OY	835	CCTGCAGAGCCAAAGCCCCCGAGAGAGAACCAAGCCCACTAGAGACTGCGGAGCTGGGA	894		
Db	491	TCTGCAGAGCAAAAGCCCCCGAGAGACACCAAACCCGCTAGAGAGCTGCGTGGGA	550		
OY	895	GAAAGACAGGAGCCCCCGCTCAGGTTCAAGGTTGACAGCCGCTGAGCACTGTGCTG	954		
Db	551	GAAAGACAGGAGCCCCCACTCAGGTTCAAGGTTGACAGCCACCTGGAGCAGCAGACTG	610		
OY	955	CTACCAACATTTGCTAGATGAGAACATGAGAGAGAGAAACAATTAATTGATCTGCCGG	1014		
Db	611	CTATCAACCAATGCTGTGATGAGAACATTGAGAGAGAGAAACAATTAATTGATCTGCCGG	670		
OY	1015	GATAGAAAATCAACAGTCCAAATTTGGGCTGGGCTCCCTTCGAGGCCAGAAAGTACGA	1074		
Db	671	GATAGAAATCAACAGTCTCAAGTTTGAACGGGATCCCTTCGAGGCCAGAAAGTACGA	730		
OY	1075	ACTGCCCCCGCACTTCAATCCACTGATCTTGCTCCATGAGCCGGAAGCATCA	1133		
Db	731	GCTGCCCCCTGAAATCTCAACCCCACTGCTGTGCTCCACAGAGCCAGAAAGTGCACA	790		
OY	1135	CATCCCAACCGAAAGCCCAAGCGTGGACCCGGGCTCTTTCCACTTCTTTGACACCC	1199		
Db	791	CATCCCAACCCGAGGCCCCCAAGGTGGAGCCAGGCTCTTTCCACTTCTTTGACACCC	850		

QY 1195 AATGCCAAGCTCTCAGAGCTCCAGCAACGAGCTCCGGGGCAACCCAGAGCGGAGCAACGA 1254
 DB 851 ATTTGGCAAGGATCAGAGCTCCAGCAACGAGCTCCGGGGCACTCAGATGGAGCAAGCA 910
 QY 1255 CTTTGTGAGGTCCCGCAAGGCCAGGCGCAAGGTGTGGGTGGGCGACGTGGCGCAAGG 1314
 DB 911 CTTTGTGAGGTCCCGCAAGGCCAGGCGCAAGCATGTGTGGGCGCAAGGCGCGCAAGG 970
 QY 1315 GGCAGAAACAGAGCCCTCTGTGAGACCGGCATCCCTGTGGGTGTCCCTCCGCGCACT 1374
 DB 971 TGCAGAGAGAGAGCTCTCAGTGTACCAACCACTCATCTGTCTCCCTCTGTGCCATCT 1030
 QY 1375 GGCCTGCCAGCCCGGCTCTCTCCCTCTCTAGCCCTCTGGGCGCAAGAGCAACGA 1434
 DB 1031 GGCACACAGCCAGCCCTCTCTCCCACTCTGGCAACCTCTGGGCGCAAGAGCAACGA 1090
 QY 1435 CCGAGCCAAAGAGAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1494
 DB 1091 TCGAGCCAAAGAGAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1150
 QY 1495 ---CCCTGTCTGTGGGCGGAGCACTGTGGGAGCTGCGGCTGTGGGTGTATGAG 1551
 DB 1151 CTCCACCGTATGGGGGGGAGAGCGTGAAGGAGCTGCTGCGTGTGTATGAGAG 1210
 QY 1552 CCAAGCCGGGAGCGGCTCAAGAGCATGAGCACCAACCAACCAACCAACCAACCA 1611
 DB 1211 CCAAGCGTGGGAGCGGCTCAAGAGCATGAGCACCAACCAACCAACCAACCAACCA 1270
 QY 1612 TTACCAACCATTTCTACAGCATGAGACCTCCCTCCGAGGGCGGCGGCGGCGGCGGCGG 1671
 DB 1271 TTACCAACCATTTCTACAGCATGAGACCTCCCTCCGAGGGCGGCGGCGGCGGCGG 1319
 QY 1672 GACCCCAACCCCGGAGCAACAGAGCATTTATTTATTTATTTATTTATTTATTTAT 1731
 DB 1320 GACCCAGCC---ACACCTTAAGGCAATTTATTTATTTATTTATTTATTTATTTAT 1376
 QY 1732 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1791
 DB 1377 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1436
 QY 1792 CTATGAAACAGCAAGCACTTTATTT 1821
 DB 1437 ACATGAAACAGCAAGCACTTTATTT 1466

RESULT 5
 AY412098 885 bp DNA linear GSS 16-DEC-2003
 LOCUS Homo sapiens NKD1 gene, VIRUTAL TRANSCRIPT, partial sequence,
 ACCESSION AY412098
 VERSION AY412098.1 GI:39768063
 KEYWORDS GSS.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 885)
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarimal,A.,
 Todd,M.A., Tanenbaum,D.M., Clevello,D.R., Lu,F., Murphy,B.,
 Ferritera,S., Wang,G., Zheng,X.H., White,T.J., Snihsy,J.J.,
 Adams,M.D. and Cargill,M.
 Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 Science 302 (5652), 1960-1963 (2003)
 JOURNAL PUBMED 14671302
 REFERENCE 2 (bases 1 to 885)
 AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarimal,A.,
 Todd,M.A., Tanenbaum,D.M., Clevello,D.R., Lu,F., Murphy,B.,
 Ferritera,S., Wang,G., Zheng,X.H., White,T.J., Snihsy,J.J.,
 Adams,M.D. and Cargill,M.
 TITLE Direct Submission

JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 COMMENT This sequence was made by sequencing genomic exons and ordering
 them based on alignment.
 FEATURES location/Qualifiers
 source 1..885
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 <1..2885
 /gene="NKD1"
 /locus_tag="HCM444"

ORIGIN
 Query Match 47.6%; Score 885; DB 9; Length 885;
 Best Local Similarity 100.0%; Pred. No. 3.1e-179;
 Matches 885; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 416 GAGCTGTGGGCGGAGCGGTTGAGAGACAGCTCAGGAGGAGAGAGAGAGAGAGAGAGAGAG 475
 DB 1 GAGCTGTGGGCGGAGCGGTTGAGAGACAGCTCAGGAGGAGAGAGAGAGAGAGAGAGAGAG 60
 QY 476 CTGGAAGTGCCCTGCTCTCTGAGAAAGTGAAGGCTGGGCGAGCGGAGATGAGAGAG 535
 DB 61 CTGGAAGTGCCCTGCTCTCTGAGAAAGTGAAGGCTGGGCGAGCGGAGATGAGAGAG 120
 QY 536 ATGAGAGAGTGAAGGAGACCTTGCCAGGCTCCAGAGAGAGAGAGAGAGAGAGAGAG 595
 DB 121 ATGAGAGAGTGAAGGAGACCTTGCCAGGCTCCAGAGAGAGAGAGAGAGAGAGAGAG 180
 QY 596 CAGTGGCACTGTTCATGAG 655
 DB 181 CAGTGGCACTGTTCATGAG 240
 QY 656 TTTGACAACAACGAGCAAGGTCAACCCGAGAGAGATCAACGATTGTGACACCATAT 715
 DB 241 TTTGACAACAACGAGCAAGGTCAACCCGAGAGAGATCAACGATTGTGACACCATAT 300
 QY 716 GAGGTGTGAGCTCTCTGTGAACAACCTGCCAATCCAGCAAGATGTGGGTTAAG 775
 DB 301 GAGGTGTGAGCTCTCTGTGAACAACCTGCCAATCCAGCAAGATGTGGGTTAAG 360
 QY 776 CTACCGTGGCCCCCATAGGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 835
 DB 361 CTACCGTGGCCCCCATAGGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
 QY 836 CTGAGAGCGCAAGGCCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 895
 DB 421 CTGAGAGCGCAAGGCCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
 QY 896 AAGAAGCAGGAGCCCGGCTCAGGTTCCAGGTTGACAGCCGCTGAGAGAGTGGCTGC 955
 DB 481 AAGAAGCAGGAGCCCGGCTCAGGTTCCAGGTTGACAGCCGCTGAGAGAGTGGCTGC 540
 QY 956 TACCAACATTCGTATGAG 1015
 DB 541 TACCAACATTCGTATGAG 600
 QY 1016 ATGAAATATCAAGTCCCAATTTGGGCTGTGCTCCCTTCGTTGGGCCAGAGAGTCA 1075
 DB 601 ATGAAATATCAAGTCCCAATTTGGGCTGTGCTCCCTTCGTTGGGCCAGAGAGTCA 660
 QY 1076 CTGCCCCCGGAGCTTCAGATCCCACTGATCTGCTGCCATGAGCCGGAAGCCATCCAC 1135
 DB 661 CTGCCCCCGGAGCTTCAGATCCCACTGATCTGCTGCCATGAGCCGGAAGCCATCCAC 720
 QY 1136 ATCCCAACCGAAAGCCCGCAAGGCGTGAACCGGCTCTCTTCAACCTTCAACCCCA 1195
 DB 721 ATCCCAACCGAAAGCCCGCAAGGCGTGAACCGGCTCTCTTCAACCTTCAACCCCA 780
 QY 1196 ATGCGCAAGGCTCAAGAGCTCAAGCAACGAGCTCCGGGCGACCCAGAGCGGAGCAAC 1255
 DB 781 ATGCGCAAGGCTCAAGAGCTCAAGCAACGAGCTCCGGGCGACCCAGAGCGGAGCAAC 840

QY 1256 TTTGTGAGTCCCGCAGGCGCAGAGTGTGGTGGG 1300
DB 841 TTTGTGAGTCCCGCAGGCGCAGAGTGTGGTGGG 885

RESULT 6
B0645656
LOCUS B0645656
DEFINITION B0645656 921 bp mRNA linear EST 15-JUN-2002
AGENCOURT_8355700 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6285268
5', mRNA sequence.
ACCESSION B0645656
VERSION B0645656.1 GI:21769828
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 921)
NIH-MGC http://mgi.nci.nih.gov/.
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
http://image.llnl.gov
Plate: LNCM2484 row: d column: 05
High quality sequence start: 16
High quality sequence stop: 685.
Location/Qualifiers
1..921
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6285268"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_100"
/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCACGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

ORIGIN
Query Match 44.5%; Score 826.8; DB 5; Length 921;
Best Local Similarity 97.6%; Pred. No. 9.8e-167;
Matches 861; Conservative 0; Mismatches 17; Indels 4; Gaps 2;

QY 1065 AGAAGTCAGAACTGCCCCCGGACCTCCATCCACTGATCTGCTCCCATGAGCGG 1124
DB 263 AGAAGTCAGAACTGCCCCCGGACCTCCATCCACTGATCTGCTCCCATGAGCGG 322

QY 1125 AAGCATCCACATCCCAACCGAAGCCCGAGGCTGACCCCGGCTCTTCCACTTCC 1184
DB 323 AAGCATCCACATCCCAACCGAAGCCCGAAGGCTGACCCCGGCTCTTCCACTTCC 382

QY 1185 TTGACACCCCAATTCGCAAGTCTCAGAGTCCAGAAAGGCTCCGGGGAGCCAGAGAG 1244
DB 383 TTGACACCCCAATTCGCAAGTCTCAGAGTCCAGAAAGGCTCCGGGGAGCCAGAGAG 442

QY 1245 GGAAGACACTTTGTAGAGTCCCGCAAGGCCAGGAGCAAGATGTGTGTGGCCACG 1304
DB 443 GGAAGACACTTTGTAGAGTCCCGCAAGGCCAGGAGCAAGATGTGTGTGGCCACG 502

QY 1305 TGGCCAGAGGGGCAAGAAACAGACCCCTGTGAGACCCGCAATCCCTGCGTGTCCCT 1364
DB 503 TGGCCAGAGGGGCAAGAAACAGACCCCTGTGAGACCCGCAATCCCTGCGTGTCCCT 562

QY 1365 CCGGCCACTGTGCTGCAAGCCCGGCTCTCTCCCTCCCTAGGCCCCCTGGGCAAGAG 1424
DB 563 CCGGCCACTGTGCTGCAAGCCCGGCTCTCTCCCTCCCTAGGCCCCCTGGGCAAGAG 622

QY 1425 AGCACAAGCACGAGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1484
DB 623 AGCACAAGCACGAGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 682

QY 1485 CTTGAGGTGGCCCTGTCTCTGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1544
DB 683 CTTGAGGTGGCCCTGTCTCTGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 742

QY 1545 ATGAGAGCCAGGCGCGGAGCGGCTCCAGAGCATGAGGACCAACCATGAAATATC 1604
DB 743 ATGAGAGCCAGGCGCGGAGCGGCTCCAGAGCATGAGGACCAACCATGAAATATC 802

QY 1605 ACCACCATTCACACACTTTCACAGACATAT--GAGGCTCTCCGAGGAGGAGGAGGAGGAG 1661
DB 803 ACCACCATTCACACACTTTCACAGACATATATATATATATATATATATATATATATAT 862

QY 1662 -CCATATGAAGACCCCGGAGCAGCAGCAAGGCAATAT 1702
DB 863 CCATATGAAGACCCCGGAGCAGCAGCAAGGCAATAT 904

RESULT 7
B0653673
LOCUS B0653673
DEFINITION B0653673 931 bp mRNA linear EST 15-JUN-2002
AGENCOURT_8419128 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6284223
5', mRNA sequence.
ACCESSION B0653673
VERSION B0653673.1 GI:21777845
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 931)
NIH-MGC http://mgi.nci.nih.gov/.
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
http://image.llnl.gov
Plate: LNCM2481 row: h column: 16
High quality sequence stop: 656.

Db	61	TGCGGAGCTGGGAGAAAGACGACGCCGCCGCTCAGGTTCCAGGTTCAACGCCGCTGG	120
Qy	942	AGCAGTCTGGCTGTCTACCAACAATTGCGTAGATGAGAACATCGAGAGAAACCACTACT	1001
Db	121	AGCAGTCTGGCTGTCTACCAACAATTGCGTAGATGAGAACATCGAGAGAAACCACTACT	180
Qy	1002	TAGATCTGGCGGGGATTAGAAACTACAGTCCCAATTTGGGCTGGGCTCCCTTCCGG	1061
Db	181	TAGATCTGGCGGGGATTAGAAACTACAGTCCCAATTTGGGCTGGGCTCCCTTCCGG	240
Qy	1062	CCCAAGATCAGAACTGCCCCCCCCGCACTCCCAATCCCACTCGATCTGCTCCCATGAGC	1121
Db	241	CCCAAGATCAGAACTGCCCCCCCCGCACTCCCAATCCCACTCGATCTGCTCCCATGAGC	300
Qy	1122	CGAAGCCATCCACATCCCAACCGAAAAGCCCAAGGCGTGACCCGGGCTCTTCCACT	1181
Db	301	CGAAGCCATCCACATCCCAACCGAAAAGCCCAAGGCGTGACCCGGGCTCTTCCACT	360
Qy	1182	TGCTTGAACACCCCAATCGCCAAAGTCTCAGAGTCTCAGCAACGGTCCGGGGACCCACG	1241
Db	361	TGCTTGAACACCCCAATCGCCAAAGTCTCAGAGTCTCAGCAACGGTCCGGGGACCCACG	420
Qy	1242	ACGGAGCAGACACTTGTGAGGTCCCCCAAGGCCCCAGGAGCAAGAGTGGGTGGGAGC	1301
Db	421	ACGGAGCAGACACTTGTGAGGTCCCCCAAGGCCCCAGGAGCAAGAGTGGGTGGGAGC	480
Qy	1302	ACGTGAGCAGAGGGGCAAGAAACAAGCCCCCTTGAGAACCGGCATCTCGGTGTGCC	1361
Db	481	ACGTGAGCAGAGGGGCAAGAAACAAGCCCCCTTGAGAACCGGCATCTCGGTGTGCC	540
Qy	1362	CCTCGGCCCACTGGCTGCGACGCCGGGCTCTCTCCCTTACGCCCTCCGGGACA	1421
Db	541	CCTCGGCCCACTGGCTGCGACGCCGGGCTCTCTCCCTTACGCCCTCCGGGACA	600
Qy	1422	AGAGACAACAAGCAGAGCGAGCAAGAGAGAGCAGAGGCTGCGGGGGCTGACAGGACAC	1481
Db	601	AGAGACAACAAGCAGAGCGAGCAAGAGAGAGCAGAGGCTGCGGGGGCTGACAGGACAC	660
Qy	1482	TGGCCTCAGGTGGCCCTGTCTGTGGGGCGGAGCACTCGGGAGCTGCGCGCTTGTGG	1541
Db	661	TGGCCTCAGGTGGCCCTGTCTGTGGGGCGGAGCACTCGGGAGCTGCGCGCTTGTGG	720
Qy	1542	TGTATGAGAGCCAGGCG-GGGCAAGCCGGTCCAGAGACATGAGCACACACACCATGAA	1600
Db	721	TGTATGAGAGCCAGGCGCGGGGAGCGCGTCCAGAGACATGAGCACACACACCATGAA	780
Qy	1601	CATGACACACATTACACCA-CTTCTACACAGCATAGAGCCCC--TCCCAGGGGCCCCAC	1657
Db	781	CATGACACACATTACACCACTTCTACAGCATAGAGCCCCCTCCCAGGGGCCCCAC	840
Qy	1658	CCTGCATATGAGG----ACCCCAAGCCCCGACACACAAGGCATTATATTTATTAAT	1713
Db	841	CCTGCATATGAGAGAACCCACCCCGGACACACACAGGGGATTATTTATTCAT	900
Qy	1714	TATTT 1719	
Db	901	TATTT 906	
RESULT 9			
Bo644956			
LOCUS	Bo644956	928 bp	mRNA
DEFINITION	AGENCOURT 8511562 NIH_MGC_100 Homo sapiens cdna clone IMAGE:62970322	linear	EST 15-JUL-2002
ACCESSION	Bo644956		
VERSION	Bo644956		
KEYWORDS	Bo644956.1 GI:21769128		
SOURCE	EST.		
ORGANISM	Homo sapiens (human)		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Euteria; Primates; Catarrhini; Homidae; Homo.		
REFERENCE	1 (bases 1 to 928)		
AUTHORS	NIH-MGC http://mgc.mci.nih.gov/ .		

TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: CGAP (Stanford) cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LMC2505 row: n column: 09 High quality sequence start: 17 High quality sequence stop: 551.
FEATURES	Location/Qualifiers
source	1..928 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:6297032" /tissue_type="hepatocellular carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /clone_id="NIH_MGC_100" /name="Organ: liver; Vector: POTB1; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."
ORIGIN	
Query Match	42.7%; Score 793; DB 5; Length 928;
Best Local Similarity	97.0%; Pred. No. 1,8e-159;
Matches	850; Conservative 0; Mismatches 21; Indels 5; Gaps 4;
Db	825 ATCAGGCTGACTGTGAGAGCGCAAGGCCCGGAGAGAGACCACCACTGAGGACTGC 884
Qy	19 ATCAGGCTGACTGTGAGAGCGCAAGGCCCGGAGAGAGACCACCACTGAGGACTGC 78
Db	885 GGAAGCTGGGAAAGAAGAGAGGAGCCCGCTCAGATTCCAGGGTGACAGCCGCTTGAGC 944
Qy	79 GGAAGCTGGGAAAGAAGAGAGGAGCCCGCTCAGATTCCAGGGTGACAGCCGCTTGAGC 138
Db	945 AGTCGGGCTGACACCATTTGCGATGATGAGAACATCGAAGAGAGAAAACAATACTTAG 1004
Qy	139 ACCTCGGCTGCTACCACTATTTGCGATGATGAGAACATCGAAGAGAGAAAACAATACTTAG 198
Db	1005 ATCTCGCGGGATGAAAACTACAGCTCCCAATTTGGGCTTGAGTCCCTTCGCTGAGCC 1064
Qy	199 ATCTCGCGGGATGAAAACTACAGCTCCCAATTTGGGCTTGAGTCCCTTCGCTGAGCC 258
Db	1065 AGAAGTCGAAACTGCCCCCGGACATTCACATTCGATCTCGTCCCATAGACCGG 1124
Qy	259 AGAAGTCGAAACTGCCCCCGGACATTCACATTCGATCTCGTCCCATAGACCGG 318
Db	1125 AAGCCATCACATCCCACACCGGAAAGCCCCGAGGCGTGGACCGGCGCTCCCTCACTTC 1184
Qy	319 AAGCCATCACATCCCACACCGGAAAGCCCCGAGGCGTGGACCGGCGCTCCCTCACTTC 378
Db	1185 TTGACACCCCATGCGCCAAAGGCTTCAGAGCTTCAGCAACGGCTTCGGGGGACCCAGAG 1244
Qy	379 TTGACACCCCATGCGCCAAAGGCTTCAGAGCTTCAGCAACGGCTTCGGGGGACCCAGAG 438
Db	1245 GGAGCAGACACTTTGTGAGTTCCTCCCAAGGCCCAAGGCGAAGAGTGTGGGTGGCCACG 1304
Qy	439 GGAGCAGACACTTTGTGAGTTCCTCCCAAGGCCCAAGGCGAAGAGTGTGGGTGGCCACG 498
Db	1305 TGGCCAGAGGGGCAAGAAACAAGCCCTCTGGGACCCGCGCATCTCGGGGTGTCCTCCCT 1364
Qy	499 TGGCCAGAGGGGCAAGAAACAAGCCCTCTGGGACCCGCGCATCTCGGGGTGTCCTCCCT 558

REFERENCE 1 (bases 1 to 953)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: sgabds-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LLCM2784 row: d column: 18
 High quality sequence stop: 627.
 Location/Qualifiers
 1..953
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6580002"
 /issue_type="teratocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_id="NIH_MGC_109"
 /note="Organ: ovary; Vector: pOT87; Site 1: EcoRI; Site 2: XhoI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 40.8%; Score 757.6; DB 5; Length 953;
 Best Local Similarity 97.5%; Pred. No. 7.1e-152;
 Matches 833; Conservative 0; Mismatches 14; Indels 7; Gaps 6;

QY 66 AGTCGGGCGCGGAGGAGGCGGCTCCGCGGCTCCGCTCGG 125
 |||||
 DB 66 AGTCGGGCGCGGAGGAGGCGGCTCCGCGGCTCCGCTCGG 125
 126 CTCGGGGGCTCTTCGGGAGAGAGCCAAAGGAGCGCCAGCGCGCGCGGCGG 185
 |||||
 DB 126 CTCGGGGGCTCTTCGGGAGAGAGCCAAAGGAGCGCCAGCGCGCGCGGCGG 185
 126 CTGGGGGGCTGCTTCGGGAGAGAGCCAAAGGAGCGCCAGCGCGCGCGGCGG 185
 |||||
 QY 186 CATGGCTTAAAGAGGCTCCCGCGCGCCAGCATGGGAAACTTCACTCCAAGC 245
 |||||
 DB 186 CATGGCTTAAAGAGGCTCCCGCGCGCCAGCATGGGAAACTTCACTCCAAGC 245
 246 CGGCGCGCGGTGCAAGGAGGAGGCGCGGAAAGGTCAGTTGGCGGTGAGCGCTG 305
 |||||
 DB 246 CGGCGCGCGGTGCAAGGAGGAGGCGCGGAAAGGTCAGTTGGCGGTGAGCGCTG 305
 306 CTTGGGCTCGGAAAGGAGATCGAGATCGGAGACAGCGCTGCGCGCGGTCT 365
 |||||
 DB 306 CTTGGGCTCGGAAAGGAGATCGAGATCGGAGACAGCGCTGCGCGCGGTCT 365
 306 CTTGGGCTCGGAAAGGAGATCGAGATCGGAGACAGCGCTGCGCGCGGTCT 365
 |||||
 QY 366 CGGAGCCCGCAACAGCTCGGTGGCGGCAACATAGGCCGGAAGCCTCGGTGG 425
 |||||
 DB 366 CGGAGCCCGCAACAGCTCGGTGGCGGCAACATAGGCCGGAAGCCTCGGTGG 425
 426 GCGAGGTGTGAGAGACAGCTCAGCGAGAGAGAGAGAGAGCTTTCGCTGGAAGTGG 485
 |||||
 DB 426 GCGAGGTGTGAGAGACAGCTCAGCGAGAGAGAGAGAGAGCTTTCGCTGGAAGTGG 485
 486 CCCTGCTCTCTGAGAAGACTGACGCGCTGGCAGCGAGATGAGAAGAGATGAGAGAG 545
 |||||
 DB 486 CCCTGCTCTCTGAGAAGACTGACGCGCTGGCAGCGAGATGAGAAGAGATGAGAGAG 545
 546 TGAGCGAACTCTGCCAGGCTTCAAAGAGCAGCTGAAGTTGAAGAGCTCCAGTGGCAGG 605
 |||||
 DB 546 TGAGCGAACTCTGCCAGGCTTCAAAGAGCAGCTGAAGTTGAAGAGCTCCAGTGGCAGG 605

QY 606 TGTTCATGAGAGAGACAGCGCGGAGAGTGAACCTTCACTTATGACTTTGACACA 665
 |||||
 DB 606 TGTTCATGAGAGAGACAGCGCGGAGAGTGAACCTTCACTTATGACTTTGACACA 665
 666 ACCGCAAGGTCACTCCGAGAGATCAACAGCTTGTGCAACACTATATAGGTGGTGG 725
 |||||
 DB 666 ACCGCAAGGTCACTCCGAGAGATCAACAGCTTGTGCAACACTATATAGGTGGTGG 725
 726 ACTTCCTTGTCAACCACTCCCAACATCCGAGAAATGCTG-CGGTAAAGCTCAACCTG 784
 |||||
 DB 726 ACTTCCTTGTCAACCACTCCCAACATCCGAGAAATGCTG-CGGTAAAGCTCAACCTG 784
 785 G-CCCCGATGAGAGAGAGAGAG-AGGAGCTCTCTTCAATCAAGCTGACCTGC-AG 841
 |||||
 DB 785 G-CCCCGATGAGAGAGAGAGAG-AGGAGCTCTCTTCAATCAAGCTGACCTGC-AG 841
 842 AGCGCAAGGCGCGGAGAGACCAAG-CCCACTGAGGACC--TGGGAGCTGGAGAG 898
 |||||
 DB 842 AGCGCAAGGCGCGGAGAGACCAAG-CCCACTGAGGACC--TGGGAGCTGGAGAG 898
 846 AGCGCAAGGCGCGGAGAGACCAAG-CCCACTGAGGACC--TGGGAGCTGGAGAG 905
 |||||
 QY 899 AAGCAGGAGACCCG 912
 |||||
 DB 906 AAGCAGGAGACCCG 919

RESULT 12
 AY412099 885 bp DNA linear GSS 16-DEC-2003
 LOCUS Pan troglodytes NKD1 gene, VIRUAL TRANSCRIPT, partial sequence,
 DEFINITION genomic survey sequence.
 ACCESSION AY412099
 VERSION AY412099.1 GI:39768064
 KEYWORDS GSS.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
 1 (bases 1 to 885)
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
 Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,
 Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smutsky,J.J.,
 Adams,M.D. and Cargill,M.
 Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 Science 302 (5652), 1960-1963 (2003)
 JOURNAL Science 302 (5652), 1960-1963 (2003)
 PUBMED 14671302
 REFERENCE 2 (bases 1 to 885)
 AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
 Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,
 Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smutsky,J.J.,
 Adams,M.D. and Cargill,M.
 Direct Submission
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 COMMENT This sequence was made by sequencing genomic exons and ordering
 them based on alignment.
 FEATURES
 source
 1..885
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 <1..>885
 /gene="NKD1"
 /locus_tag="HGM4444"

ORIGIN

Query Match 40.3%; Score 748.8; DB 9; Length 885;
 Best Local Similarity 84.7%; Pred. No. 5.4e-150;
 Matches 750; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 416 GAGCTCGTGGGCAAGTGTGAGAGACAGCTCAGCGAGAGAGAGAGAGAGCTTTGG 475
 |||||

```

Db      1 GAGCTGTGGGCGACGCTGTGAGAGACGCTCAGCGAAGGAGAGACGACTTTCGG 60
Qy      476 CTGGAAGTGGCCCTGCTCTCTGAGAACTGACGGGCTGGGCAAGCGAGATGAGAAAG 535
Db      61 CTGGAAGTGGCCCTGCTCTCTGAGAACTGACGGGCTGGGCAAGCGAGATGAGAAAG 120
Qy      536 ATGGAAGAGTGAAGGCAACCTGCGCCAGGCTCCAAAGAGAGTGAAGTTGAAGGCTC 595
Db      121 ATGGAAGAGTGAAGGCAACCTGCGCCAGGCTCCAAAGAGAGTGAAGTTGAAGGCTC 180
Qy      596 CAGTGGCAAGTGTCCATGAGAGAGAGACAGCGGCGAGAGTGAAGTCACTTCACTGTATGAC 655
Db      181 CAGTGGCAAGTGTCTCTGAGAGAGAGACAGCGGCGAGAGTGAAGTCACTTCACTGTATGAC 240
Qy      656 TTTGACAAACGCGCAAGGTCACCCGAGAGATCAACGCTTGTGTGACACCATCTAT 715
Db      241 TTTGACAAACGCGCAAGGTCACCCGAGAGATCAACGCTTGTGTGACACCATCTAT 300
Qy      716 GAGGTGTGACCTCTCTCTGCAACCACTCCCAACATCCAGCAAGATGCTGCGGTTAAAG 775
Db      301 GAGGTGTGACCTCTCTCTGCAACCACTCCCAACATCCAGCAAGATGCTGCGGTTAAAG 360
Qy      776 CTACCGTGGGCCCCCGATGAGCAGCAGAGAGAGAGCGCTTGTCAATCAGGCTGAC 835
Db      361 CTACCGTGGGCCCCCGATGAGCAGCAGAGAGAGAGCGCTTGTCAATCAGGCTGAC 420
Qy      836 CTGCAAGAGCGCAAGGCCCCGAGAGAGAGACCAAGCCCACTGAGAGACTGCGGAGTGGAG 895
Db      421 CTGCAAGAGCGCAAGGCCCCGAGAGAGAGACCAAGCCCACTGAGAGACTGCGGAGTGGAG 480
Qy      896 AAGAAGCAGGAGGCCCCGCTCAGGTTCCAGGTTGACAGCGGCTGAGAGAGTCTGCTGC 955
Db      481 AAGAAGCAGGAGGCCCCGCTCAGGTTCCAGGTTGACAGCGGCTGAGAGAGTCTGCTGC 540
Qy      956 TACCACTGTGGTGTAGTGAAGCATCGAGAGAGAAACCACTAGTCTGCGCGG 1015
Db      541 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 600
Qy      1016 ATGAAATACACGCTCCCAATTTGGGCTCGGCTTCCGTTGCGGCCAGAAAGTCAGAA 1075
Db      601 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 660
Qy      1076 CTGCCCCCGCGCACTTCATTCGACTCGATCTGCTCCATGAGCGCGGAAGCATCTAC 1135
Db      661 CTGCCCCCGCGCACTTCATTCGACTCGATCTGCTCCATGAGCGCGGAAGCATCTAC 720
Qy      1136 ATCCCAACCGGAAGGCCCCGAGGCGTGGAGCCGCGCTCTTCCACTTCTTGAACCCCA 1195
Db      721 GTCCCAACCGGAAGGCCCCGAGGCGTGGAGCCGCGCTCTTCCACTTCTTGAACCCCA 780
Qy      1196 ATGCGCAAGGCTTCAGAGCTTCAGCAACGCTCCGAGGCAAGCGAGAGAGAGAGC 1255
Db      781 ATGCGCAAGGCTTCAGAGCTTCAGCAACGCTCCGAGGCAAGCGAGAGAGAGAGC 840
Qy      1256 TTTGTGAGTGTCCCAAGGCGCCAGGCGCAAGAGTGTGGGTGGGCT 1300
Db      841 TTTGTGAGTGTCCCAAGGCGCCAGGCGCAAGAGTGTGGGTGGGCT 885

RESULT 13
LOCUS   B0649813
DEFINITION AGENCOURT 8302499 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6271026
5', mRNA sequence.
ACCESSION B0649813
VERSION   B0649813.1
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 966)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/

```

```

TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL  Unpublished (1999)
CONTACT  Robert Strausberg, Ph.D.
EMAIL    cgabbs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Stanford Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2447 row: b column: 19
High quality sequence stop: 583.

FEATURES
Source
1..966
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6271026"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_100"
/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the Laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

ORIGIN
Query Match      38.3%; Score 712.6; DB 5; Length 966;
Best Local Similarity 94.6%; Pred. No. 3.3e-142;
Matches 759; Conservative 0; Mismatches 40; Indels 3; Gaps 2;

Qy      825 ATAGAGCTGACCTGCAAGAGCGCAAGGCCCGAGCAGAGACCAAGCCCACTGAGAGCTGC 884
Db      3 ATAGAGCTGACCTGCAAGAGCGCAAGGCCCGAGCAGAGACCAAGCCCACTGAGAGCTGC 62
Qy      885 GGAAGTGGGAGAAAGAGCGAGCCCGCTCAGGTTCCAGGTTGACAGCGGCTGGAGC 944
Db      63 GGAAGTGGGAGAAAGAGCGAGCCCGCTCAGGTTCCAGGTTGACAGCGGCTGGAGC 122
Qy      945 AGTCTGGCTGTACCACTTGGTATGAGAAATCGAGAGAGAAACCACTACTTAG 1004
Db      123 AGTCTGGCTGTACCACTTGGTATGAGAAATCGAGAGAGAAACCACTACTTAG 182
Qy      1005 ATCTGCGCGGATGAAACTACAGTCCCAATTTGGGCTTGGCTCCCTTCGTGGCCC 1064
Db      183 ATCTGCGCGGATGAAACTACAGTCCCAATTTGGGCTTGGCTCCCTTCGTGGCCC 242
Qy      1065 AGAAGTCAGAACTGCCCCCGACCTCCAAATCCCACTGATCTGCTCCATGAGCCGG 1124
Db      243 AGAAGTCAGAACTGCCCCCGACCTCCAAATCCCACTGATCTGCTCCATGAGCCGG 302
Qy      1125 AAGCCATCCACATCCCAACCGAAAGCCCAAGGCTGAGACCGGCTCTTCCACTTCC 1184
Db      303 AAGCCATCCACATCCCAACCGAAAGCCCAAGGCTGAGACCGGCTCTTCCACTTCC 362
Qy      1185 TTGACACCCCAATGCGCAAGGTTCTCAGAGTCCAGCAACGCTCCGAGGCAAGCCAGAGC 1244
Db      363 TTGACACCCCAATGCGCAAGGTTCTCAGAGTCCAGCAACGCTCCGAGGCAAGCCAGAGC 422
Qy      1245 GGAAGAGCACTTGTGAGGTCTCCCAAGGCGCAAGGCGCAAGAGTGTGGGTGGGCGACG 1304
Db      423 GGAAGAGCACTTGTGAGGTCTCCCAAGGCGCAAGGCGCAAGAGTGTGGGTGGGCGACG 482
Qy      1305 TGCCAGAGGGGAGAAACAGCCCTCTGAGAACCGGCATCTGCGGTGTCCCT 1364
Db      483 TGCCAGAGGGGAGAAACAGCCCTCTGAGAACCGGCATCTGCGGTGTCCCT 542

```

Oy		1265	CGGGCCACCTGAGCTGCAGAGCCCGCCTTCCTCCCTTAAGCCCCCTTGGGGCAACA	1424
Db		543	CGGCCACTTGATGCTGCAGAGCCCGGAGCTCTCTCCCTCTAAGCCCCCTTGGGGCAACA	602
Oy		1425	AGCAAAGCACCGAACCCAAAGAGAAGCCAGCAGAGGCTGTCCGGGGCTGTGACGACCATCTGG	1484
Db		603	AGCAAAGCACCGAACCCAAAGAGAAGCCAGAGAGGCTGTCCGGGGCTGTGACGACCATCTGG	662
Oy		1485	CCTCAGGTGGCCCCCTGTCTCTGGGGGCGGGAGCACCCTG--CGGAGACTGCCCTTGTGTGT	1542
Db		663	CCTCAGGTGGCCCCCTGTCTCTGGGGGCGGGAGCACCCTG--CGGAGACTGCCCTTGTGTGT	722
Oy		1543	GTAATGAGAGCCAGAGCCGGGGCAAGCC--GGTCCAGAGA CATGAACACACACACATGATGAC	1601
Db		723	GTAATGACACCAAGGTGTGGGCTCCCGGGGTCCAGAGACATTCAACCCCACCATTCAGAGGAC	782
Oy		1602	ATCACCAACCATTTACACCACTT	1623
Db		783	GTCCCCCCCCGTTCCTCCCGGCGGT	804
RESULT 14				
LOCUS	BQ652087	996 bp	mRNA	linear EST 15-JUN-2002
DEFINITION	AGENCOURT_8207495 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6283110			
ACCESSION	BQ652087			
VERSION	BQ652087.1	GI:21776259		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE	1 (bases 1 to 996)			
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.			
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL	unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs@emall.nih.gov Tissue Procurement: CGAP (Stanford) cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov Plate: LILCM2478 row: j column: 07 High quality sequence stop: 604.			
FEATURES				
Source				
	1..996			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:6283110"			
	/tissue_type="hepatocellular carcinoma, cell line"			
	/lab_host="DH10B (phage-resistant)"			
	/clone_id="NIH_MGC_100"			
	/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."			
ORIGIN				
	Query Match	38.3%	Score 711.4;	DB 5; Length 996;
	Best Local Similarity	95.6%;	Pred. No. 5.9e-142;	
	Matches 743;	Conservative 0;	Mismatches 31;	Indels 3; Gaps 1;
Oy	822 TCAATCAGGCTGACCTGCAGAGCGCCGACGAGAGACCAAGCCCATGANGACC	881		

[illegible]

DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: LNCM2514 row: a column: 21
 High quality sequence start: 17
 High quality sequence stop: 553.

FEATURES

source

1..1008

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAG:6300188"

/issue_type="hepatocellular carcinoma, cell line"

/lab_host="NIH MGC 100"

/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:

EcoRI; cDNA made by oligo-dT priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor:

GGGCGAG(G). Size-selected >500bp for average insert size

1.8kb. Library constructed by Ling Hong in the Laboratory

of Gerald M. Rubin (University of California, Berkeley)

using ZAP-cDNA synthesis kit (Stratagene) and Superscript

II RT (Life Technologies). Note: this is a NIH_MGC

Library."

ORIGIN

Query Match 38.0%; Score 706; DB 5; Length 1008;

Best Local Similarity 92.5%; Pred. No. 8.5e-141;

Matches 835; Conservative 0; Mismatches 46; Indels 22; Gaps 8;

QY 825 ATCAGGCTGACCTGCGAGAGCGCAAGGCCCGAGCAAGACCAAGCCACTGAGACCTGC 884
 DB 23 ATCAGGCTGACCTGCGAGAGCGCAAGGCCCGAGCAAGACCAAGCCACTGAGACCTGC 82
 QY 885 GGAGCTGGGAGAGAGAGCGAGCGCCCGCTCAGGTTCCAGGGTGAAGCGCGCTGGAGC 944
 DB 83 GAGAGCTGGGAGAGAGAGCGAGCGCCCGCTCAGGTTCCAGGGTGAAGCGCGCTGGAGC 142
 QY 945 AGTCTGGCTGCTACCACTTGGCTAGTAGAATCGAGAGAGAAACCACTACTTAG 1004
 DB 143 AGTCTGGCTGCTACCACTTGGCTAGTAGAATCGAGAGAGAAACCACTACTTAG 202
 QY 1005 ATCTCCCGGAGATGAAAACTACAGTCCCAATTTGGGCTTCCCTTCCGTGGCCC 1064
 DB 203 ATCTCCCGGAGATGAAAACTACAGTCCCAATTTGGGCTTCCCTTCCGTGGCCC 262
 QY 1065 AGAAGTCAGAACTGCCCCCGGCACTCCAAATCCCACTGATTCGCTCCATGAGCCGG 1124
 DB 263 AGAAGTCAGAACTGCCCCCGGCACTCCAAATCCCACTGATTCGCTCCATGAGCCGG 322
 QY 1125 AAGCCATCCATCCCAACCGAAAGCCCAAGGCGTGAACCGGCTCTTCCACTTCC 1184
 DB 323 AAGCCATCCATCCCAACCGAAAGCCCAAGGCGTGAACCGGCTCTTCCACTTCC 382
 QY 1185 TTGACACCCCAATCGCCAAAGTCTCAGAGTCCAGCAACGCGCTCCGAGGCAACCAGAGC 1244
 DB 383 TTGACACCCCAATCGCCAAAGTCTCAGAGTCCAGCAACGCGCTCCGAGGCAACCAGAGC 442
 QY 1245 GGAGCAAGCACTTTGAGAGTCCCGCAAGGCCCAAGGCAAGAGTGTGGTGTGGGCACG 1304
 DB 443 GGAGCAAGCACTTTGAGAGTCCCGCAAGGCCCAAGGCAAGAGTGTGGTGTGGGCACG 502
 QY 1305 TGGCCAGAGGGGCAAGAAACAAGCCCCCTTGTGGGACCCGCGCAT-CCCTGGAGTGTCCCC 1363
 DB 503 TGGCCAGAGGGGCAAGAAACAAGCCCCCTTGTGGGACCCGCGCATCCCTGGAGTGTCCCC 562
 QY 1364 TCCGCCACCTGGCTGCGAGCCCGGCTCTCCCTCCCTAGCCCCCTTGGGCACAAG 1423
 DB 563 TCCGCCACCTGGCTGCGAGCCCGGCTCTCCCTCCCTAGCCCCCTTGGGCACAAG 622
 QY 1424 AAGCAACAAGCAGGCGCAAGAGGAGCGAGGGGTGCGGGGGCTTGACAGGACCACTG 1483

DB 623 AAGCAACAAGCAGGCGCAAGAGGAGCGAGGCTGCGGGGGCTTGACAGGACCACTG 682
 QY 1484 GCCTCAGGTGGCCCTG-TCCCTGGGGCGGAGACCC-TGCGGAGCTGCCCC-CTTG 1537
 DB 683 GCCTCAGGTGGCCCTGTTCTTGGGGCGGAGGACCTTGCGGAGCTGCCCCGCGCTTGGG 742
 QY 1538 GTGCTATGAGAGCCAGGCC--GGGCAAGCGGTCCAGAGACATGAGACCACTCA---C 1591
 DB 743 GGGGGGTATGAAAGCCCGGAGCCGAGGCGAGCGGGTCAAGAAACATGAGACCACTCACC 802
 QY 1592 CACCATGAACATACACCATTTACCACTTCTACCAAGATAGAGCCCC--TCCCC 1647
 DB 803 CACCATGAACATACACCATTTACCACTTCTACCAAGATAGAGCCCCCTTCCCC 862
 QY 1648 AGGGCCCCACCTGACATA----TGAAGAGCCCAACCCCGGACACCAAGGCAATATT 1702
 DB 863 AGGGGCCAAACCCGGGCAAAATGAAAGAAACCAACCCCGGAAACCCCAAGGGTTATT 922
 QY 1703 ATT 1705
 DB 923 AAT 925

Search completed: December 29, 2004, 20:34:18
 Job time : 8801 secs

THIS PAGE IS BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 30, 2004, 03:23:44 ; Search time 141 Seconds
(without alignments)
2369.296 Million cell updates/sec

Title: US-09-993-966-7
Perfect score: 2499
Sequence: 1 MGLKSKPAVACVKRSPSG.....RHHHHHHHHHHHHFYQT 470

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 35534441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xip
-Q/cgnt2 1/USPTO.spool.p/US09993966/runat.29122004.130759.9030/app.query.fasta.1.647
-DB=Issued Patents NA -QPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPTC=0
-LOOPTXT=0 -INITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdt
-LIST=45 -DOCLIN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09993966 @CGN 1.128 @runat.29122004.130759.9030 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MAN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Issued Patents NA.*

1: /cgnt2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgnt2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgnt2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgnt2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgnt2_6/ptodata/1/ina/PTCUS.COMB.seq.*
6: /cgnt2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2499	99.8	1438	4	US-09-506-066E-5
2	2174.5	87.0	1731	4	US-09-506-066E-3
3	783	31.3	1285	4	US-09-506-066E-7
4	583.5	23.3	1307	4	US-09-506-066E-9
5	215.5	8.6	4954	4	US-09-506-066E-1
6	160.5	6.4	2063	4	US-09-270-767-14537
7	133	6.1	1449	4	US-09-252-991A-15134
8	153	6.1	2226	4	US-09-252-991A-14877
9	152	6.1	2268	4	US-09-252-991A-14877
10	152	6.1	2433	4	US-09-999-689A-2
11	152	6.1	2433	4	US-09-944-807-3
12	149.5	6.0	4403765	3	US-09-103-840A-2

13	149.5	6.0	4411529	3	US-09-103-840A-1	Sequence 1, Appl1
14	146.5	5.9	2576	1	US-08-471-033-35	Sequence 35, Appl1
15	146.5	5.9	2576	2	US-08-471-044-35	Sequence 35, Appl1
16	146.5	5.9	2576	2	US-08-463-483A-35	Sequence 35, Appl1
17	146.5	5.9	2576	2	US-08-471-046A-35	Sequence 35, Appl1
18	146.5	5.9	2576	2	US-08-470-566B-35	Sequence 35, Appl1
19	146.5	5.9	2576	2	US-08-469-334-35	Sequence 35, Appl1
20	146.5	5.9	2576	3	US-09-300-529-35	Sequence 35, Appl1
21	146.5	5.9	2655	1	US-08-471-033-17	Sequence 17, Appl1
22	146.5	5.9	2655	1	US-08-471-033-26	Sequence 26, Appl1
23	146.5	5.9	2655	2	US-08-471-044-17	Sequence 17, Appl1
24	146.5	5.9	2655	2	US-08-471-044-26	Sequence 26, Appl1
25	146.5	5.9	2655	2	US-08-463-483A-17	Sequence 17, Appl1
26	146.5	5.9	2655	2	US-08-463-483A-26	Sequence 26, Appl1
27	146.5	5.9	2655	2	US-08-471-046A-17	Sequence 17, Appl1
28	146.5	5.9	2655	2	US-08-471-046A-26	Sequence 26, Appl1
29	146.5	5.9	2655	2	US-08-470-566B-17	Sequence 17, Appl1
30	146.5	5.9	2655	2	US-08-470-566B-26	Sequence 26, Appl1
31	146.5	5.9	2655	2	US-08-469-334-17	Sequence 17, Appl1
32	146.5	5.9	2655	2	US-08-469-334-26	Sequence 26, Appl1
33	146.5	5.9	2655	3	US-09-300-529-17	Sequence 17, Appl1
34	146.5	5.9	2655	3	US-09-300-529-26	Sequence 26, Appl1
35	146.5	5.9	4031	1	US-08-471-033-49	Sequence 49, Appl1
36	146.5	5.9	4031	2	US-08-471-044-49	Sequence 49, Appl1
37	146.5	5.9	4031	2	US-08-463-483A-49	Sequence 49, Appl1
38	146.5	5.9	4031	2	US-08-471-046A-49	Sequence 49, Appl1
39	146.5	5.9	4031	2	US-08-470-566B-49	Sequence 49, Appl1
40	146.5	5.9	4031	2	US-08-469-334-49	Sequence 49, Appl1
41	146.5	5.9	4031	3	US-09-300-529-49	Sequence 49, Appl1
42	146.5	5.8	5194	1	US-09-599-652-1	Sequence 1, Appl1
43	146	5.8	5194	2	US-08-642-846-1	Sequence 1, Appl1
44	146	5.8	5194	4	US-09-264-604-1	Sequence 1, Appl1
45	146	5.8	5194	4	US-10-002-389-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-09-506-066E-5
Sequence 5, Application US/09506066E
Patent No. 6630323
GENERAL INFORMATION:
APPLICANT: Scott, Matthew
APPLICANT: Wharton, Keith
TITLE OF INVENTION: Naked Cuticle Genes and their Uses
FILE REFERENCE: STAN-121
CURRENT APPLICATION NUMBER: US/09/506,066E
CURRENT FILING DATE: 2000-02-17
PRIOR APPLICATION NUMBER: 60/120,646
PRIOR FILING DATE: 1999-02-17
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 1438
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (8)...(1418)
OTHER INFORMATION: Mkd1 coding sequence
US-09-506-066E-5
Alignment Scores:
Pred. No.: 8.39e-206
Score: 2493.00
Percent Similarity: 99.79%
Best Local Similarity: 99.79%
Query Match: 99.76%
DB: 4
Gaps: 0
US-09-993-966-7 (1-470) x US-09-506-066E-5 (1-1438)

```

QY 1 MercGlyLysLeuHisSerLysProAlaAlaValCysIleYsArgArgLysSerProGluGly 20
Db 8 ATGGGGAAATTCCTCACTCAAGCCGGCCGGCTGTGGCAAGCAGAGAGAGCCGGAAGGT 67
QY 21 AspSerPheAlaValSerAlaAlaATPAlaArgLysGlyIleGluGluTrpIleGlyArg 40
Db 68 GACAGCTTCGCGCGTGAAGCGCTGCGGCTCGAAGGAGCATCGAGGTGGATCGGAGGA 127
QY 41 GluArgCysProGlyGlyValSerGlyProArgGluLeuArgLeuAlaGlyThrIleGly 60
Db 128 CAGGCGTCCCGCGCGGTCTCGGAGCCCGCAGCTGCGGTGGCGGAGCATGAGC 187
QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80
Db 188 CGAAGCACCGGGAGCTGTGGGAGCGTGTGGAGACAGCTCAGCGAGAGAGAGAG 247
QY 81 AspArgPheArgLeuGluValAlaLeuProGluTrpIleArgGlyLeuGlySerGly 100
Db 248 GACGACTTTCGGCTGGAAGTGGCTGCTCTGAGAAAGCTGACGGGCTGGCAGCGGA 307
QY 101 AspGlyLysLysMetGluArgValSerGluProCysProGlySerLysGluLeuLys 120
Db 308 GATGAGAAAGAGATGAGAGAGTGAAGCACTCTGCCAGGCTTCAAGAAAGCAGCTGAG 367
QY 121 PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGlnGluTrpThrPhe 140
Db 368 TTTCAGAAAGCTCCAGTCCAGCTGTCCATGGAGAGAGACAGCCGGCAGAGTGAACCTTC 427
QY 141 ThrLeuTyraPheAspAsnAsnGlyLysValThrArgIleAspIleThrSerLeuLeu 160
Db 428 ACCGTATGACTTTGACAAACAGGAGGTGACCCGAGAGAGCATCAGCAGCTTGCTG 487
QY 161 HisThrIleTyrgluValAlaAspSerSerValAsnHisSerProThrSerSerLysMet 180
Db 488 CACCACTATATGAGGTGGTGACTCTGTCTCAACCACTCCCAATCCAGAAAGATG 547
QY 181 LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysValSerValLeuVal 200
Db 548 CTGGGGGTAAAGCTCACCGTGGCCCGATGGCAGCCAGAGCAAGAGAGGCTCTTGTC 607
QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgIleGluThrLysProThrGluAspLeu 220
Db 608 AATCAGGCTGACTGACAGAGCGCAAGGCCCGAGCAGAGCAAGCCACATGAGAGACTG 667
QY 221 ArgSerTrpGluLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
Db 668 CGGAGCTGGAGAGAAAGCAAGCGCCCGCTCAGGTTCAGGGGTGACAGCCGCTGAG 727
QY 241 GlnSerGlyCysTyraHisCysValAspGluAsnIleGluArgArgAsnHisTyrlLeu 260
Db 728 CAGCTGGCTGCTCAACCATTTGGCTGATGAGAACATCGAGAGAGAAACCACTACTTA 787
QY 261 AspLeuAlaGlyIleGluAsnTyraThrSerGlnPheGlyProGlySerProSerValAla 280
Db 788 GATCTCGCGGGATGAAAACCTACACGTCCTGAGCTGGCTCCCTTCCTGAGGCC 847
QY 281 GlnLysSerGluLeuProArgThrSerAsnProThrArgSerArgSerHisGluPro 300
Db 848 CAGAGGTCAAGAACTGCCCCCGCAGCTTCATATCCATCTCATCTCCCAAGAACCG 907
QY 301 GluAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPhe 320
Db 908 GAAACCATCCACATCCACAGCCGAAAGCCCAAGGCTGAGCCGGGCTCTTCACATTC 967
QY 321 LeuAspThrProIleAlaLysValSerGluLeuGlnArgLeuArgGlyThrGlnAsp 340
Db 968 CTTGACACCCCAATCGCAAGGTCTCAGAGCTCCAGAGACGGCTCGGGGACATCAGAGAC 1027
QY 341 GlySerLysHisPheValArgSerProLysValAlaGlnGlyLysSerValGlyValHis 360
Db 1028 GGGGAGCAGGCACTTGTGGAGGTCCCCCAAGGCCAGGAGCAAGATGTGGGTGGGCAC 1087
QY 361 ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380

```

```

Db 1088 GTGGCCAGAGGGGCAAGAAACAGCCCTCTGGAGACCGGCATTCCTGGGTGTCCTCC 1147
QY 381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
Db 1148 TCCGCCACCTGGCTGCTCAGCCCGGCTCTCTCCCTCCCTTACGCCCTCCGAGCACAG 1207
QY 401 LysHisLysHisArgArgAlaLysGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
Db 1208 AAGCACAAGCACGAGCAGAGAGAGCAGCAGAGCTGCGGGAGCTTGACAGCACTG 1267
QY 421 AlaSerGlyGlyProValLeuGlyArgGluHisLeuArgGluLeuProAlaLeuVal 440
Db 1268 GCCTCAGTGCGCTGTCTGCGGGCGGAGACCTGGGAGAGCTGCCCTTGTTGTGTG 1327
QY 441 TyrgluSerGlnAlaGlyGlnProValGlnArgHisGluHisHisHisHisHisHis 460
Db 1328 TATGAGAGCCAGGCGCGGCTCAGAGACATGAGACACACACCATGAACAT 1387
QY 461 HisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 470
Db 1388 CACCACCATTAACACCATTTCTACAGACA 1417

RESULT 2
US-09-506-066E-3
; Sequence 3: Application US/09506066E
; Patent No. 6630323
; GENERAL INFORMATION:
; APPLICANT: Scott, Matthew
; APPLICANT: Wharton, Keith
; APPLICANT: Zeng, Wenlin
; TITLE OR INVENTION: Naked Cuticle Genes and their Uses
; FILE REFERENCE: STRA-121
; CURRENT APPLICATION NUMBER: US/09/506,066E
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/120,646
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (140)...(1553)
; OTHER INFORMATION: Nkdl coding sequence
US-09-506-066E-3

Alignment Scores:
Pred. No.: 3,46E-178 Length: 1731
Score: 2174.50 Matches: 407
Percent Similarity: 90.64% Conservative: 19
Best Local Similarity: 86.60% Mismatches: 43
Query Match: 87.01% Indels: 1
DB: 4 Gaps: 1

US-09-993-966-7 (1-470) x US-09-506-066E-3 (1-1731)
QY 1 MercGlyLysLeuHisSerLysProAlaAlaValCysIleYsArgArgLysSerProGluGly 20
Db 140 ATGGGGAAATTCCTCACTCAAGCCGGCCGGCTGTGGCAAGCAGAGAGAGCCGGAAGGT 199
QY 21 AspSerPheAlaValSerAlaAlaATPAlaArgLysGlyIleGluGluTrpIleGlyArg 40
Db 200 GACAGCTTTCGGCTGGAAGTGGCTGCTCTGAGAAAGCTGACAGGTGAGATCGGAGG 259
QY 41 GluArgCysProGlyGlyValSerGlyProArgGluLeuArgLeuAlaGlyThrIleGly 60
Db 260 CAGGCGTTCGCGCGGTGAAGCGCTGCGGCTCGAAGGAGCAAGGTGAGATCGGAGG 319
QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80

```

Db	330	CGAGGCACTCGGGAAGCTCGTGGGTGACACTTCTAGAGAGGCTCTCGGTGAGGAGCGAG	379
Qy	81	AspApPheAaVgLeuGluValAlaLeuProPArgLysThrAspGlyLeuGlySerGly	100
Db	380	GACGACTTCCCCCTTGAAGTGGCCCTGCGCTGAGAAATGACAGAGCTTAGTGTGGA	439
Qy	101	AspGlyLysMetGlyIleArgValSerGluProCysProGlySerLysIleLeuLys	120
Db	440	GATGAGAAAGAAATGAGAGACTGAGACCAACTGGCCAGGCTTCCAGAAAGAGCTCAAG	499
Qy	121	PheGluGluLeuGlnCysAspValSerMetGluIleAspSerArgGlnIleThrPhe	140
Db	500	TTTGAAAGAGCTACAGTGTGATCTCTCTGTGAGAGAGAACAGCCGGCAAGTGGACTTTC	559
Qy	141	ThrLeuThrAspPheAspMetGlyLysValThrArgLysAspIleThrSerLeuLeu	160
Db	560	ACTCATATATGACTTCCACAAACATGGCAAAAGTGAACCCGTGAGAGAACTTCCAGCTTGGCTG	619
Qy	161	HisThrIleTyrgLysValValAspSerSerValAspHisSerProThrSerSerLysMet	180
Db	620	CATACCACTTATGAAGTGGTGTACTCTCTGTGAACCAATTTCCCACTACAGAGAACAGA	679
Qy	181	LeuArgValIleLeuThrValAlaProAspGlySerGlnSerLysArgSerValLeuVal	200
Db	680	CTGCGGGGGAAGCTCACCGTGGCTCTTGACGGGAGCGAGAGTAAAGAGAGCTCTTTTC	739
Qy	201	AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeu	220
Db	740	AACCATACCGGATCTGCAGAGCACAAAGCCCCGAGACAGACAAACCCGCTGAGGAGCTG	799
Qy	221	ArgSerTrpGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu	240
Db	800	CGTGGCTGGAGAAAGACGAGGAGACCCCACTCAGGTTCAGAGGTGACAGCACTGGAG	859
Qy	241	GlnSerGlyCysTyrgHisIleCysValAspGluAsnIleGluIleArgArgAsnIleTyrgLeu	260
Db	860	CAGCCAGAGCTGTACCACTTGGCTGATGAGAACTTGAAGAGAGAAACATACCTACCTA	919
Qy	261	AspLeuAlaGlyIleGluAsnTyrgThrSerGlnPheGlyProGlySerProSerValAla	280
Db	920	GACCTGGCGGGATGAGAACTACAGCTCAGTTTGGACCGGATCCCTTCGGTGGCC	979
Qy	281	GlnLysSerGluLeuProProArgThrSerAsnProThrArgSerThrGSerHisGluPro	300
Db	980	CAGAAGTGAGAGCTGCCCTCCGTAATCTCCAAACCCCACTCGCTTCGCTCCACAGAGCA	1039
Qy	301	GluAlaIleHisIleProHisIleArgLysProGlnGlyValAspProAlaSerPheHisPhe	320
Db	1040	GAACTGCCCACTATCCACACCGGAGCCCCCAAGGTTGAGACCAAGCTCTTCCACTTC	1099
Qy	321	LeuAspThrProIleAlaLysValSerGluLeuGlnArgLeuArgGlyThrGlnAsp	340
Db	1100	CTTGACACCCCATTTGGCCAAAGCACTAAGCTCCAGCAACGGCTCCGGGCACTCAGAT	1159
Qy	341	GlySerLysHisPheValArgSerProLysValGlnGlyLysSerValGlyValGlyHis	360
Db	1160	GGGAGCAAGCACTTGTGAGGTGCCCCCAAGGCCCAAGGCAAGAAACATGGTGTGGGCCAC	1219
Qy	361	ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro	380
Db	1220	GGGGCGCAAGGGGCAAGAAAGCAAGGCTCCACTGTGATCCCAACCAACCCATCTGTCTCCCC	1279
Qy	381	SerAlaHisIleLeuAlaAspProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys	400
Db	1280	TCTGCCCATCTGGCCACAGCCCAACCCCTTCTCCCAACCTTGCAACCTTGGGGGCAAG	1339
Qy	401	LysHisLysHisIleArgAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu	420
Db	1340	AAACCAAGAGCTCGAGCAAGAGAGCCAGCGAGACTGCGGGGCTCGAGAGGCCCCCTG	1399
Qy	421	AlaSerGlyGlyProValLeuGlnGlyArgGlyHisIleAspArgLysLeuProAlaLeuVal	439
Db	1400	GCTGGAGGAGGCTCCACCGTCACTTGGGGCGGAGCAGATGAGAACTGCTGCGCTGTAGT	1459

QY 440 ValTyrGluSerGlnAlaGlyIleProValGlnAspGluGluIleHisHisHisGluIle 459

Db 1460 GTGTACGAGAGCCAGCTTAGCAGGCCGCTCCAGAACACGAAACCATCACCACCCGAA 151

QY 460 HisHisHisHisIleTyrHisHisIlePheTyrGln 469

Db 1520 CATCCACACCATATATACCACTTCTATACG 1549

RESULT 3
US-09-506-066E-7

; Sequence 7, Application US/09506066E
; Patent No. 6630323
; GENERAL INFORMATION.

GENERAL INFORMATION:
; APPLICANT: Scott, Matthew
; APPLICANT: Wharton, Keith

```

; APPLICANT: Zeng, Wenlin
; TITLE OF INVENTION: Naked Cuticle Genes and their Uses
; FILE REFERENCE: STAN-121

```

; CURRENT APPLICATION NUMBER: US/09/506,066EE
 ; CURRENT FILING DATE: 2000-02-17
 ; PRIOR APPLICATION NUMBER: 60/120,646

; PRIOR FILING DATE: 1999-02-17
 ; NUMBER OF SEQ ID NOS: 15

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1285

```

```

; GENOMES: 1200
;
; TYPE: DNA
; ORGANISM: Mus musculus

```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)...(1282)

```

OTHER INFORMATION: Nkd2 coding sequence
US-09-506-066E-7

Alignment Scores
Pred. No.:

Score:	783.00	Matches:	180
Percent Similarity:	56.86%	Conservative:	52
Best Local Similarity:	44.12%	Mismatches:	134

Query Match:	31.33%	Indels:	42
DB:	4	Gaps:	10

US-09-993-966-7 (1-470) x US-09-506-066E-7 (1-1285)

QY 80 GIUaspapbPheargLeuGIuValAlaLeuProProGIuLysThrAspGly----- 96

QY 97 -----LeuGIysSerGIyAaGIlulysIlySmeGluIuaGValSerGIuProCys 112

Db 164 CAGGGCCAGCTCTTCAGCACAGATGTGTTGGGAGAAAGGCAGCAGCCGTGAGGGTCCACTG 22

QY 113 ProGlySerIysLysGlnLeuLysPheGluGluLeuGlnCysAspValSerMetGluGlu 133

133 AspSerArgGlnGluTyrPheThrPheThrLeuTyrAspPheAspAsnGlnGlyValThr 155

Db 284 GACACCCCGAGATGACATTGATGACTTGACACAGTGGGAAATCACC 343

```

153 ArgGluApIleIeRhSerLeuLeuHISThrIeTyrgIuValAlaSpSerValAsn 172
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```

DB 344 AGGAGGACATGTCAGCCGATGACACCATCTACGAGGTGGTCGATGCCCTGTCAT 403

Db 404 CACTCTCCGGCAGCAGACCTTTCAGTGAAGCTAACTGTACGCCCTGAACCTCC 463

Oy 193 GInserlyBArgeSValleuValnGlnlAlAspluEugInSerAlarGProArGAlA 212

Db 464 AGCAAGAGGAAATGTCCTTCACCTGGCCAA---GACCGGAGGCCACTCGTGGCAGACAA 530

```

Qy 213 GluThrlYsProthrlGluAspLeuArgSerTrpGluYsGlnAlaAlaProLeuArg 232
Db 521 GAGATTGAGCTCAGATGAGAGCCCGGATGGCTGTGACAGAGGCTATTCCGGCTTACAGCAG 580
Qy 233 PheGlnGluYasPseArgLeuGlnGlnSerGlyCysTyrHisCysValAspGluAsn 252
Db 581 AAGCCCAATGCTGATCCCAAGCCCTGTCTGTGCGAGTCCCTTACTGTGTGGATGAGAAC 640
Qy 253 IlleGluArgArgAsnHisTyrLeuAspLeuAlaGlyIlleGluAsnTyrThrSerGlnPhe 272
Db 641 ACAAGAGCGCAAAACCACTACTAGACTCTTGTGGCATCGAGAACTACATCACTTAAGTT 700
Qy 273 GlyProGlySerProSerValAlaGlnLysSerGluLeuProAlaGlyThrSerAsn-- 291
Db 701 GGTCCTGGGTACACCACTGACAGGCGCAGGCAAGAACATCATGGCAGGGCCACACATTT 760
Qy 292 ProHisArgSerArgAsnHisGlnProGlnAlaIleHisIleProHisArgLysProGln 311
Db 761 CCAAGCAGGTCCCATTCACAAAGTCGGATGTCCACGGCTATACCAACCGCAGGTCTCA 820
Qy 312 -----GlyValAspProAlaSerPheHisPheLeuAspThrProIleAlaLys 327
Db 821 GTCTCGGTGAGCATGTCATACACAGT-----AAGACGCTGTCCACCCGG 865
Qy 328 ValSerGluLeuGlnGlnAlaGluLeuArgGlyThrGlnAspGlySerLysPheValArg 347
Db 866 GCCCTGCTGCACAGCCCGGATGACAGGGCAGAG-----AAGCAGTTCCTCAG 916
Qy 348 SerProLysAlaGlnGlyLysSerValGlyValGlnHisValAlaArgGlyValArgAsn 367
Db 917 TCTCTTAAGGCTCCAGAAACCTCTTGAGACA-----CAAGCAGTGGC 961
Qy 368 LysProProLeuGlyProAla-----IleProAlaVal 378
Db 962 AAGCCA-----GGGAAGCTCTCAGATTCCTGCGCAGGCCGCTGCTATGCCCCAGAGT 1015
Qy 379 SerProSerAlaHisPheLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGly 398
Db 1016 GCTCAGATGAGGCACACACTTCTCAGCCCCCACCACAGCCTCCACCGCAGCCCTTAGT 1075
Qy 399 HisValYsHisPheHisArgAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAla 418
Db 1076 CACAAAGCGGTACCGGCAGAAAGCCAGAAAGGCCACTCACCATTAAAGGCGATGGCCAG 1135
Qy 419 ProLeuAlaSerGlyGlyProValLeuGlnLysArgGlnHisPheLysGluLeuProAlaLeu 438
Db 1136 CCT-----ACCATGTGAGCATGAAAGTAAAGTACCGGACTGCTGCTCCATG 1180
Qy 439 ValValTyrGluSerGlnAlaGlyGlnProValGlnArgHisGlnHisPheHisPheHis 458
Db 1181 CTGGGGCGCTGAGGGCTATGTATGTGCTGTGGTCCAGAGGATGAACACACATACCAT 1240
Qy 459 GlnHisHisHisHisTyrHisHis 466
Db 1241 GAGCACACACACATCACCACCAC 1264

RESULT 4
US-09-506-066E-9
; Sequence 9, Application US/09506066E
; Patent No. 6630323
; GENERAL INFORMATION:
; APPLICANT: Scott, Matthew
; APPLICANT: Wharton, Keith
; APPLICANT: Zeng, Wenlin
; TITLE OF INVENTION: Naked Cuticle Genes and their Uses
; FILE REFERENCE: STAN-121
; CURRENT APPLICATION NUMBER: US/09/506,066E
; CURRENT FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/120,646
; PRIOR FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9

```

	LENGTH: 1307		
/	TYPE: DNA		
/	ORGANISM: Homo sapiens		
US-09-506-066E-9			
Alignment Scores:			
Pred. No.:	4,95e-41	Length:	1307
Score:	583.50	Matches:	144
Percent Similarity:	54.25%	Conservative:	41
Best Local Similarity:	42.23%	Mismatches:	103
Query Match:	23.35%	Indels:	53
DB:	4	Gaps:	10
US-09-993-966-7 (1-470) x US-09-506-066E-9 (1-1307)			
QY	150	LYSVATHTRAGLUASPILETHRSERLEUENHISTHRIETRYGLVALVALHSPSER	169
DB	52	GAGGTACACAGGAGAGACATGTCCAGCTCATGACACACATATATAGAGTCTGTGATGCC	111
QY	170	SERYALAENHISSEPROTHRSERSELYSWETLEUATGVALLYLEUTHVALAPRO	189
DB	112	TCGGTCAACCACTCTCCGGGACAGCAAGACCTTCCTGGTGAAGCTAACCGTCAACCTCT	171
QY	130	ASPGLYSERGINSERLYHARGSERVALLEUVALAENGINALASPUENLINSERLAARG	209
DB	172	GAGCCCTCAGACAGAGAGAGAGAGGAGTCTCTGTGCGCAGAGACCGGAGCCACCGGT	231
QY	210	PROAGTLAGLUTHRELYEPROTHRGLUASPLEUATGSEETTPGLULYSGLINARGALA	229
DB	232	TGCAGAGATGAGAGGTGMACTGGCAAGAGCCAGAGGCTGTCAGACAGAGTTGTGTGCA	291
QY	230	PROLEUATGPLEGINGLYASPSERATGLEUGINGLINSERGLYCYTRYHISHICYVAL	249
DB	292	CACGTACAGAGAGCCAGTACTGACCCCGACCCCTGCTCGGACGGGGGCGCTTACTCGTG	351
QY	250	ASPGIUAENILLEGUATGTHGASNHISTYTRLEUASPLEUVALGLYILEGLUANTYRTTHR	269
DB	352	GACGAGAAACACGAGACGCGAGAAACCATACCTGCGAGCTCGCGGATTGAGAACTACAG	411
QY	270	SERGINPHEGLPROGLYSERPROSESERVALAAGLINSERGLUENUPROPROATGTHR	289
DB	412	TTCAGATTTGGGCTTGGGCTCCCTCTCTGTGCAAGCAAGACAGAGGCCACAGGCGAGGCC	471
QY	250	SERAEEN---PROTHRARGSERATGSEERHISGLUPROGLUALATLEHISILEPROHISARG	308
DB	472	TCGCAACCTCCAGGCCCGGCTCCGCTCCAGAGCCAGACATGCGTACACACACCGC	531
QY	309	LYSPROGLH-----GLYVALASPPROGLASERPHENISHELEUASPTHRPRO	324
DB	532	AGGTACACAGGTGCTGTGGAGACACGTGCGGCCAGCTCG-----GAGCCT	576
QY	335	ILEALALYVALSERGLULENGINGLINSERGLYUASERGLYTHRGINASPGLYSERLYSHIS	344
DB	577	GCTGCCCCGGGCTCTGACACGACGCCCGCGGAGAGGGGCCGAG-----AGCAG	627
QY	345	PHEVALARGSERPROLYALAGLINSGLYUASERVALGLYVALGLYHISVALALARGLY	364
DB	628	TTCTTCACAGTCCCCCAAGGCGTCCGGGAAGCGC-----	660
QY	365	ALARGAENLYSPROPROLEUGLYPROALALPEPROALVALASERPROSER-----ALA	382
DB	661	-----CCTGGGGTGGCCAGCAGCAGCAAGTCCGGGAAAGCC	696
QY	363	HISLEUVALALASERPROALALEUENUPROSERLEUVALA-----	395
DB	697	TTACAGCTACTACTCGCGGCTGTGCGCGCCCGCCAGGCGCCCTTCAGAGCGGCCACCACTTC	756
QY	396	-----PROLEUGLINSGLYSHISLEUASERGLYVALYSGLINSER	409
DB	757	CCGACAGCCCCCAGCCGACCTACGCGCCAGAGCGGTACCGCCAAAAGGCGAGGAG---	813
QY	410	GLINGLINSGLYASRGGLYLEUGINALAPROLEUVALASERGLYGLYPROVALLEUGLYARG	429

```

Db      814 -----GGCACTCGGCACTCAAGCCGCCACGCTCAGCCTGCACAGTGT----- 858
QY      430 GlnHis-----LeuArgGluLeuProAlaLeuValValTyrGluSerGlnAlaGlyGln 447
Db      859 GACACGAGAGTGTGTGGGACCTGCGCGCCACGACGACGACGAGAGGGCTACGGGTGCCA 918
QY      448 ProValGlnArgHisGlnHisHisHisHisGlnHisHisHisHisHisHisHisHisHisHis 467
Db      919 GTGATCCAGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 978
QY      468 Tyr 468
Db      979 CAC 981

RESULT 5
US-09-506-066E-1
; Sequence 1, Application US/09506066E
; Patent No. 6630323
; GENERAL INFORMATION:
; APPLICANT: Scott, Matthew
; APPLICANT: Wharton, Keith
; APPLICANT: Zeng, Wenlin
; TITLE OF INVENTION: Naked Cuticle Genes and their Uses
; FILE REFERENCE: STAN-121
; CURRENT APPLICATION NUMBER: US/09/506,066E
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/120,646
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4954
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (966)...(3752)
US-09-506-066E-1

Alignment Scores:
Pred. No.:      2,15e-08      Length:      4954
Score:          215.50      Matches:      129
Percent Similarity: 33.00%      Conservative: 69
Best Local Similarity: 21.50%      Mismatches: 187
Query Match:      8.62%      Indels:      216
                        Gaps:      26

US-09-993-966-7 (1-470) x US-09-506-066E-1 (1-4954)
QY      41 GlnArgCysProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIle--- 59
Db      1281 CAGGATGTGAGCAGCGCGCGCCATAGCAGCATCTGCGCATCAGCAGCACTTCCAT 1340
QY      60 -----GlyArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSer 76
Db      1341 GGCACGACGCGCAAAATCAAT-----ATGACGACGACCACTGCC 1382
QY      77 GlnGlnGlnGlnuAspAspPheArgLeuGlnValAlaLeuProProGlnuLysThrAspGly 96
Db      1383 CAGAGTAGAGAT-----GTGGTGGATCGCGCTCCCAAGATCGACGACGACGAC 1433
QY      97 LeuGlySerGlyAspGlyLysMetGlnuArgValSerGluProCysProGlySerLys 116
Db      1434 ACTGGCCACGCCCACTCGCGCCACCTGCACACCAACGAG----- 1475
QY      117 LysGlnLeuLysPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 133
Db      1476 GACGCGATCGGATCGGAGGAAATTCACCTGAGCGTGCCTGGAGGCGGCAAGTATCG 1535
QY      134 SerArgGlnGlnuLysPheThrLeuLysPheAsp---AsnAsnGlyLysValThr 152
Db      1536 CACCGCGTGAAGTTCTGTTCACTTCACTGAGCGGAGCATACGCGCAAGATACA 1595

```

```

QY      153 ArgGlnAspIleThrSerLeuLeuHisThrIleTyrGlnuValAspSerSerValAsn 172
Db      1596 AAGGACGACATCTGGGCACTGTGTACACCATTCAGATTCAGATTCAGATTCAGATTCAG 1655
QY      173 HisSerProThrSerSerLysMetLeuArgValLysLeuThrValAlaProAspGlySer 192
Db      1656 GTCCCCCACTGGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1715
QY      193 GlnSerLysArg----- 196
Db      1716 TCGAAATCGACCGGTGTGCTCCGTTCCGTGACGCGGATTCAGACGACGACGCC 1775
QY      196 ----- 196
Db      1776 AGCAAACTGAAGAATTGCCCGGCTGTGGCGCATGTCCGAACCCCTGGCGGCGGA 1835
QY      197 -----SerValLeuValAsnGlnAlaAspLeuGlnSerAla--- 208
Db      1836 GAGTGGGATCGGCGGAGCGTGGCGCTAACGATCCGCGGCAACCGCGGACGAT 1895
QY      209 -----ArgProArgAlaGlnuLys---ProThrGlnuAspLeuArgSerTrpGlnuLys 225
Db      1896 CGCTATCGACGACGAACTGATTAGTCCGATGACGAGGACGATGACAGCAACAGCGAA 1955
QY      226 LysGlnArg-----AlaProLeuArgPheGln----- 234
Db      1956 AAGGAGAGAGCCCGCCCGCCCGCTTCCGCGGACGACGCGGAGAGTGAACAAAG 2015
QY      235 -----GlyAspSer-----ArgLeuGlnGlnSerGlyCysTyrHisHis----- 247
Db      2016 GCGACTGGAGAGGCGCATTCACGACGACGACGACGACGACGACGACGACGACGACGAC 2075
QY      248 -----CysValAspGlnuAsnIleGlnuArgAsnHis-----Tyr 259
Db      2076 TCCCGGCGGAGAGTGTCTGACGAGCAAGAAATACCGCCGCAATGCGCAATACCTAC 2135
QY      260 LeuAspLeuAlaGlyIleGlnuAsnTyrThrSerGlnPheGlyProGlySerProSerVal 279
Db      2136 GAGATATGCTGTAATCTCAATCTCAGTCTGACGCGGAGT---GAGCCAGGTGACGTCTC 2194
QY      280 AlaGlnLysSer---GluLeuProProArgThrSerAspProThrArgSerArgSerHis 298
Db      2195 GCACCGACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 2254
QY      299 GluPro-----GluAlaIleHisIlePro--- 306
Db      2255 GAAACAGGCGACCGCGGTCAAGATGTTGCAAGGCGCGCAACCAAAAGATCCAGGA 2314
QY      307 -----HisArgLysProGln 311
Db      2315 CCACTGCTCGAAACCGGACGCGGACGCGCTGTAGTGGGCAAGATTCGCGCTGCGGAA 2374
QY      312 GlyValAspProAlaSerPheHisPheLeuAspThrProIleAlaLysValSerGluLeu 331
Db      2375 TCGCCATCTGACCTGACGAGCGCGCGGTGGTACCCGACGCCGACGCGGTGAACCA 2434
QY      332 GlnGlnArgLeuArgGlyThrGlnAspGlySerLysHisPheValArgSerProLysVal 351
Db      2435 CAAGACGCG----- 2443
QY      352 GlnGlyLysSerValGlyValGlnHisValAlaArgGlyVala-----ArgAsnLysPro 369
Db      2444 -----GTGGAGTCCACCGCTGGCGGTGGGGGTGGGTGGGTGGGTGGGTGGGTGG 2491
QY      370 ProLeuGlyProAlaIleProAlaValSerProSerAlaHisIleuAlaAlaSerProAla 389
Db      2492 CGATGGGCTGACCTGCTGACGCGGACCGGACCCATTCCTCAACCCGAGGACGATCAACA 2551
QY      390 LeuLeuProSerLeuAlaProLeuGlnHisLys-----Lys 401
Db      2552 GCAAAATCAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 2611
QY      402 HisLysHisArgAlaLysGlnuSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 420

```

Db 2612 GAATCGCAAGATTATATGATCATCATGAGAGAGATGAGAAACCGCTGTGTTT 2671
QY 421 -----AlaSerGlyGly----- 424
Db 2672 TCAGCTGATGAGAAACCCAGCCAAATGTGATGCCATACCGGCAACCGGAGCAACA 2731
QY 425 -----ProValLeu-----GlyArgGlyHisLeuArgGlyLeu 435
Db 2732 ACAACCAACAACAAGAGAGAGCCAGCCAGCTGCATATCCGGCTCGAATAATCCAC- 2785
QY 436 ProAlaLeuValValTyrGlnSerGlnAlaGlyLeuProVal----- 449
Db 2786 -----GTTAATTATCAACACCAAGTCCGGTCCGACAGAGTCCGCGTCAGCTCAG 2836
QY 449 ----- 449
Db 2837 TCCACCCAGCGAGCCACACCCCGACATTCAGAGCTCCGCGGCATCGAGT 2896
QY 450 GlnArg-HisGlnHisHisHisGlnHis-----HisHisTyrHisHis 466
Db 2897 CAAGGTCAGACAGACCAACCCACTCATCCACTCATCCAGCCACCAACACACAC 2954

RESULT 6

US-09-270-767-14537
Sequence 14537, Application US/09270767
Patent No. 6703493
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 14537
LENGTH: 2063
TYPE: DNA
ORGANISM: Drosophila melanogaster
US-09-270-767-14537

Alignment Scores:

Pred. No.:	0.000322	Length:	2063
Score:	160.50	Matches:	81
Percent Similarity:	37.62%	Conservative:	36
Best Local Similarity:	26.05%	Mismatches:	97
Query Match:	6.42%	Indels:	99
DB:	4	Gaps:	14

US-09-993-966-7 (1-470) x US-09-270-767-14537 (1-2063)

QY 122 GlnGluLeuGlnCyAspValSerMetGlu-----GluAspSerArgGlnGluTyr 138
Db 1142 CAGAAATTCACCTCGACGTGTCCGTGAGGCGGCAAGTCATCGACCGCTCATGTC 1201
QY 139 ThrPheThrLeuTyrAspPheAsp---AsnAsnGlyValValThrArgGluAspIleThr 157
Db 1202 TCGTTCAGTTCAGACCTGAGACGGGATCAACGCAAGATTAACAAAGAGACATCGTG 1261
QY 158 SerLeuHisThrIleTyrGluValValAspSerValAsnHisSerProThrSer 177
Db 1262 GCGCTTGTGACACCATATACGATCGCAAGTCGCGGTGGTGGCCCATCGCGGC 1331
QY 178 SerIysMetLeuArgValIysLeuThrValAlaProAspGlySerGlnSerIysArgSer 197
Db 1322 AGCAAGACATCAACATGCGGCTCACCCTCAGTCCGAGGCAATGGAATCGCAGCG 1381
QY 198 ValLeu----- 199
Db 1382 GTGTGTCCCTTCCGGTGGCAGCCGATTAGACAGACCAACGCGACCAAACTGAAGAG 1441
QY 199 ----- 199

Db 1442 TTGCCACGGGTCGTGGCGGCATGTGCMAAGCCTGGCCGCGAGAGTGGATCCGCG 1501
QY 200 -----ValAsnGlnAlaAspLeuGlnSerAlaArgProArgAla 212
Db 1502 GAGCTGCGGCTTAACGACATCCGCGCAACCGCGCAGATGCTATTCACACAGCAAA 1561
QY 213 GluThrIys---ProThrGluAspLeuArgSerTyrPduIlyblybGlnArg 228
Db 1562 CTGATTATGTCGATGACGAGACGATGACAGCAACAGCAAGAAAGAAAGAGACGCGCC 1621
QY 229 ---AlaProLeuArgPheGln-----GlyAspSer--- 237
Db 1622 CAGCCCTGTCGCGCCGACCGCCAGCGCAAGTGTGAACAAGCGACTGGGAAGGCAT 1691
QY 238 ---ArgLeuGlnGlnSerGlyCysTyrHisHis----- 247
Db 1682 CACCACAGTCGACGTCCGACAGGTATACCAAGAACATTCGCCGCGGAGACATGC 1741
QY 248 CysValAspGluAsnIleGluArgArgAsnHis-----TyrLeuAspLeuAlaGlyIle 265
Db 1742 TCACAGCAACAAGATACGCCCCGACATGGCCACATATCTACAGAAATATGTGAATCTC 1801
QY 266 GluAsnTyrThrSerGlnPheGlyProGlySerProSerValAla----- 280
Db 1802 AAGTGTGCAAGCCGAGAGGT-GGACCAAGTGTGACTGTCCCTCCACCGACAGCACCA 1860
QY 281 -----GlnIysSerGluLeuProProArgThrSerAsn-----ProThr 293
Db 1861 GAGCCACCCGAAACCATTAATATGCGCAGC-AGGACATTCATGAMAGCAGGCCACCGC 1919
QY 294 ArgSerArgSerHisGlnProGluAlaIleHisIleProHisArgIysProGln---Gly 312
Db 1920 GGGTCAAGATGTTGCAAGGGCGCGCA-----AACAAAGAAAGAACCCAGCAAT 1970
QY 313 ValAspProAlaSerPheHisPheLeuAspThrProIleAlaIysValSerGluLeuGln 332
Db 1971 GTGAGTCCC-----ATACGCGCAACCGGACGACACAAACACACACAA 2015

RESULT 7

US-09-252-991A-15134
Sequence 15134, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 15134
LENGTH: 1449
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15134

Alignment Scores:

Pred. No.:	0.00084	Length:	1449
Score:	153.00 <td>Matches:<td>118</td></td>	Matches: <td>118</td>	118
Percent Similarity:	32.89% <td>Conservative:</td> <td>55</td>	Conservative:	55
Best Local Similarity:	22.43% <td>Mismatches:</td> <td>189</td>	Mismatches:	189
Query Match:	6.12% <td>Indels:</td> <td>164</td>	Indels:	164
DB:	4	Gaps:	28

US-09-993-966-7 (1-470) x US-09-252-991A-15134 (1-1449)

```

QY 14 ArgArgGluserProGluGlyAaspSerPheAlaValSerAlaAlaTrrAlaArgLysGly 33
Db 16 CGTGGCGAAGACGCGAGAGGTGAGTCAAGAACAACTAGCGCTGAGAAAGCGGCAAGT 75
QY 34 IleGluIutrrIleGlyArgGlnArg-----CyPProGly 45
Db 76 GCTCGACAGACTGTTCAATGAGCTGGGCAAGCTCGGACTCTTTCAGAAAGACCTTGCA 135
QY 46 GlyValSerGly----- 49
Db 136 GGAATGCAAGAAACCGGCGTCAATGTGTCCTGACAGTCAAGGCGACATGATGAAGAT 195
QY 50 -----ProArgGlnLeuArgLysAlaGly-----ThrIleGlyArgSerThrArgGlnLeu 66
Db 196 TTCCCAACCCGATGCTTTCGAGACATGCGGTGAGCGTGTACTACAAAGACGCTTTCGACAA 255
QY 67 ValGlyAaspValLeuArgAaspThrLeuSerGluGluGlnAaspPhe---ArgLeu 85
Db 256 GTGGGGCCAACTTTCGAAAGAGCTCGGCGTGAACCCGAAACCGGTATCTCCAGCGCTGA 315
QY 86 -----GluValAlaLeu-----ProProGluLysThrAasp 95
Db 316 CGACAAAGATCAAGTCGCTGCCCGCTCCAGCAAGAAAGAGATCTCCACAGACATCCACGA 375
QY 96 GlyLeu-----GlySerGlyAaspGlnLysLysMetGlnArgValSerGluPro 111
Db 376 GGTGTACAGCCAGCCCGGCAAAATGGCGATGTGTGACTCCGTGAAGAGCATCAACCACT 435
QY 112 CyPProGlySerLysLysGlnLeuLysPheGluGlnLeuGlnCyAaspValSerMetGlu 131
Db 436 GCACATTCGAGGAGAGCTCATGCTCGACGCTGATCCGCGCATGATCCGCAACTCCGG 495
QY 132 GluAaspSerArgGlnGluIutrrThrPheThrLeuThrAaspPheAaspAaspGlyVal 151
Db 496 CCAAGATGTGGGCGAAGGACGCGAAGCAAGAAAGACACCAAGCGGTATATGCGGAAGACAC 555
QY 152 ThrArgGluAaspIleThrSerLeuLeuHisThrIleGlyValValAaspSerSerVal 171
Db 556 CTAAGGCCCG----- 564
QY 172 AaspHisSerProThrSerSerLysMetLeuArgValLysLeuThrValAlaProAasp--- 190
Db 565 ---CATCTACAGAGATGATCAATCTTTCGAAACACCAAGCGGCTTCGACCGGACAC 621
QY 191 ---GlySerGlnSerLysAaspSerValLeuValAaspGlnAlaAaspLeuGlnSerAlaArg 209
Db 622 CATGGGCAAGGTCGCGAAGCTCGGCGCT-----GATGGCGCAGAA 660
QY 210 ProArgAlaGluIutrrLysProThrArgGluAaspLeuArgSerTrpGlu----- 224
Db 661 GGCCTGAGGATATGCGGTCCGACAGCAAGACTTCGAAATGACCGCGACGCGACATGCG 720
QY 225 -----LysLysGlnArgAlaProLeuArgPheGlnGlyAasp 236
Db 721 CGTGTGCTGCGGCGAGCGAGCGGTGATGACAGACGAGTGTGAGACCGGCGACATGTG 780
QY 237 SerArgLeuGlnLeuSerGlyCyAArgLysHisCyAValAaspGlnLeuIleGlnArgArg 256
Db 781 GCGCGCTGCGAAGCAAGCAAGCA-----CGCAC 807
QY 257 AaspHisThrLeuAaspLeuAlaGlyIleGlnAaspThrSer----- 270
Db 808 GATCCCGCAAGTGTGATGAGTGGCGGCTGACCCCGCGACGCTGATCCGATCCCGCGGAT 867
QY 271 ---GlnPheGlyProGlySer-----ProSerValAlaGlnLysSerGluLeuPro 286
Db 868 CTTTCTGCTGAGACCGGAGCGGCGCCACGACGAGAACTCGGAGAGAGGTGAGCTGTA 927
QY 287 ProArgThrSerAaspProThrArgSerArgSerHisGluProGluAlaIleHisIlePro 306
Db 928 CCAAGAGCAAGCAAGCAAGTGAACGCTGCTGCA---CATCAGCATATGAGGCTCAACAGAGGC 984

```

```

QY 307 HisArgLysProGlnGlyVal---AaspProAlaSerPheHisPheLeu----- 321
Db 985 CATTCGGTCAGCATGAGACGCTGATTCGGCGCAAGACACCATCTCGGTGACCGGCAA 1044
QY 322 -----AaspThrProIleAlaLys-----ValSerGluLeu---Gln 332
Db 1045 CGTCTGCGGAGTACTGACCGACACTGTTCCGATCATGAGAACTGGGACCTCGGCGCAA 1104
QY 333 GlnArgLeuArgGlyThrGlnAaspGlySerLysHisPheValArgSerProLysValGln 352
Db 1105 GATGCTCTCATGATGCTCGCTGATGGCGGCGCGGCGGATGTACGAAACCGCGCGGCG 1164
QY 353 GlyLysSerValGlyValGlnHisValAlaArg-----GlyAlaAaspLysPro 369
Db 1165 -----CTCGGACCCAAAGACGTCGACGAACTGTGTGAAAGAACTAATCT 1209
QY 370 ProLeuGlyProAlaIleProAlaValAaspProSerAlaHisPheAlaAlaSerProAla 389
Db 1210 GCGCTGGGA-----CTCCCTGGGCGAGTTCCTGCG----- 1239
QY 390 LeuLeuProSerLeuAlaProLeuGlnHis----- 399
Db 1240 -----CTTGCGCGGTGCTCTGAGAGAAACCGGATCAAGACCGGCAACGCGCAAGCCAA 1293
QY 400 -----LysLysHisLysHisArgAlaLysGluSerGln----- 410
Db 1294 GCTGCTCGGAGAGCGCTGACGAGACCGGCAAGCTGTGACAAACAAAGTCCGC 1353
QY 411 -----GlnGlyCyAaspArgGlyLeuGln-----AlaProLeuAlaSerGly----- 423
Db 1354 GTGGCGCAAGTGGCGGACATGCAACAGCGGCGAGCCACTTCTACTGCGAGTACTG 1413
QY 424 GlyProValLeuGlyArg 429
Db 1414 GCGCCAGGCGCTGCGCGC 1431

RESULT 8
US-09-252-991A-14877
; Sequence 14877, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIORITY FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14877
; LENGTH: 2226
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-14877

Alignment Scores:
Pred. No.: 0 0016 Length: 2226
Score: 153.00 Matches: 118
Percent Similarity: 32.89% Conservative: 55
Best Local Similarity: 22.43% Mismatches: 189
Query Match: 6.12% Indels: 164
DB: 4 Gaps: 28

US-09-993-966-7 (1-470) x US-09-252-991A-14877 (1-2226)

QY 14 ArgArgGluserProGluGlyAaspSerPheAlaValSerAlaAlaTrrAlaArgLysGly 33
Db 603 CGTGGCGAAGACGCGAGAGGTGAGTCAAGAACAACTAGCGCTGAGAAAGCGGCAAGT 662
QY 34 IleGluIutrrIleGlyArgGlnArg-----CyPProGly 45

```



```

Db      663 GCTGACAGATGTTATGAGCTGCGGCAAGCTGCGGACTTCTTGAGAAAGACCTTGCA 722
Qy      46 G1yValSerGly-----|||
Db      723 GGACTGCAAGAAACCGCGCTCATGTGTCCTGACGTCAAGCGACATGATGAAGAT 782
Qy      50 -----ProArgGlnLeuArgLeuAlaGly---Thr1LeuArgSerThrArgGluLeu 66
Db      783 TTCCCAACCCGATGCTTCGACATGCGGTGAGCGGTCTACTACAGAGACGTCTTGACAA 842
Qy      67 ValG1yAspValLeuArgAspThrLeuSerGlnGluGlnuAspAspPhe---ArgLeu 85
Db      843 GTGGGGCAACTCTTGCAAGAGCTGCGGCGTGAACCGAACAACGGTATCTCCACCGCTTA 902
Qy      86 -----GluValAlaLeu-----ProProGlnuLySerThrAsp 95
Db      903 GCACAAAGATCAAGTCGTGCGCGCTCCAGCAAGAAAGATCTCCACGACATCCACGA 962
Qy      96 G1yLeu-----GlySerG1yAspG1uLybLybMetGluArgValSerGluPro 111
Db      963 GGTGTACAGCCACCGCCCGAAATGCGATGCTGACTCCGTGAAGGCATCACCAACT 1022
Qy      112 CysProGlySerLybLybGlnLeuLybPheGlnGluLeuGlnCysAspValSerMetGlu 131
Db      1023 GCACATTCGAGGACGCTCATCGTCGACGCTCGATGCGCGCATGATCCGCACTCGG 1082
Qy      132 GluAspSerArgGlnGluTrpThrPheThrLeuTyTrAspPheAspAsnAluLybVal 151
Db      1083 CCAAGTGTGGGGCGCAAGACGCGCAAGAAAGACCAACCAAGCGGTAAATGCCGGAAGCAC 1142
Qy      152 ThrArgGluAsp1LeuThrSerLeuLeuHisThr1LeuArgValValAspSerSerVal 171
Db      1143 CTACGCCCG-----|||
Qy      172 AsnHisSerProThrSerSerLybMetLeuArgValLybLeuThrValAlaProAsp--- 190
Db      1152 ---CATTCACAGAGATGATGATCACTCTGCAAGACCAAGCGGCTTCGACCGGACAC 1208
Qy      191 ---GlySerGlnSerLybArgSerValLeuValAsnGlnAlaAspLeuGlnSerAlaArg 209
Db      1209 CATGCGAGCGGTGCGCAAGCTCGGCT-----GATGGCGCAGAA 1247
Qy      210 ProArgAlaGluTrpThrLybProThrGluAspLeuArgSerTrpGlu----- 224
Db      1248 GCGCGAAGAAATACGCGCTCCACGACAAAGACTTGAAATGACCGCCGACGACCATGCG 1307
Qy      225 -----LybLybGlnArgAlaProLeuArgPheGlnG1yAsp----- 236
Db      1308 CGTGTCTGCGCGACGCGACGCGTGTGATGACACAGACGTCGAGACCGCGGACATCTG 1367
Qy      237 SerArgLeuGlnGlnSerGlyCysTyTrHisHisCysValAspGluAsn1LeuArgArg 256
Db      1368 GCGGCGCTGCAGACCAAGCA-----CGCACG 1394
Qy      257 AsnHisTrpLeuAspLeuAlaG1y1LeuAsnTrpThrSer----- 270
Db      1395 GATCCGGACTGGGTCAAGCTGCGCTGACCGCGGCAAGCGGACGTCGATACCCGGGAT 1454
Qy      271 ---GlnPheGlyProGlySer-----ProSerValAlaGlnLybSerGluLeuPro 286
Db      1455 CTTCGTGCTGACCGCGAGCGCGCGCCGACGACGCGGAACTGCGCAAGAAAGTCGAGCTGA 1514
Qy      287 ProArgThrSerAspProThrArgSerArgSerHisGluProGluAlaLeuHis1LeuPro 306
Db      1515 CCTCAAGACCAACGACCTGACCGGTCTGCA---CATAGCATCATGGGCTTACACGAGGC 1571
Qy      307 HisArgLybProGlnGlyVal---AspProAlaSerPheHisPheLeu----- 321
Db      1572 CATTCGGCTGACGATGAAAGCGCTGATCCGCGCAAGGACACCAATTCGTGACCGGCA 1631
Qy      322 -----AspThrPro1LeuAlaLyb-----ValSerGluLeu---Gln 332

```

```

Db      1632 CGTCTGCGGCACTACCTGACCGACCTGTCCCATCATGGAATGGGCACTCGGCGCA 1691
Qy      333 GlnArgLeuArg1yThrGlnAspGlySerLybHisPheValAspSerProLybAlaGln 352
Db      1692 GATGCTCTCATGCTGCTGCTGATGCGGCGGCGGACATGTAAGAACCGGCGGCGG 1751
Qy      353 G1yLybSerValG1yValG1yHisValAlaArg-----G1yAlaArgAsnLybPro 369
Db      1752 -----CTCGGCAACCAAGCAACGTCGACGACCAAGCTGTGGAAAGAACTACT 1796
Qy      370 ProLeuGlyProAla1LeuProAlaValSerProSerAlaHisLeuAla1AspProAla 389
Db      1797 GCGCTGGGA-----CTCGCTGGCGGAGTTCTGCG----- 1826
Qy      390 LeuLeuProSerLeuAlaProLeuGlnHis----- 399
Db      1827 -----CTGGCGGTGCTCTGAGAGAAACCGGAGATCAAGACCGGCAACCGCAAGGCCAA 1880
Qy      400 -----LybLybHisLybHisArgAlaLybGlnSerGln----- 410
Db      1881 GCTGCTCGCAAGCGCTGACACCAAGCCACCGGCAAGCTGTGGACAAACAAGTCGCC 1940
Qy      411 -----GlnG1yCysArgG1yLeuGln-----AlaProLeuAlaSerGly----- 423
Db      1941 GTCGCGCAAGTTCGCGGACATGCAACAACCGGCGACGACCTTACTTGCGGATGTA 2000
Qy      424 G1yProValLeuGlnArg 429
Db      2001 GGCCCAAGCCCTGACCGC 2018

RESULT 9
US-09-252-991A-14487/C
; Sequence 14487, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196, 136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14487
; LENGTH: 2268
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14487

Alignment Scores:
Pred. No.: 0.00165 Length: 2268
Score: 153.00 Matches: 118
Percent Similarity: 32.89% Conservative: 55
Best Local Similarity: 22.43% Mismatches: 189
Query Match: 6.12% Indels: 164
DB: 4 Gaps: 28

US-09-993-966-7 (1-470) x US-09-252-991A-14487 (1-2268)
Qy      14 ArgArgGluSerProGlnGluGlyAspSerPheAlaValSer1Ala1ArgGlyGly 33
Db      1657 CGTCGCAAGACGCGCAAGGTCGACGTCGAAGAAAGCACTAGCGCTGCGAAGGCGAGT 1598
Qy      34 1LeuGluTrp1LeuG1yArgGlnArg-----CysProGly 45
Db      1597 GCTGACAGATGTTATGAGCTGCGGCAAGCTGCGGACTTCTTGAGAAAGACCTTGCA 1538
Qy      46 G1yValSerGly-----|||
Db      1537 GGACTGCAAGAAACCGCGCTCATGTGTCCTGACGTCAAGCGACATGATGAAGAT 1478

```



```

QY 50 -----ProArgGlnLeuArgLeuAlaGly-----ThrIleGlyArgSerThrArgGlnLeu 66
Db 1477 TTCACACCCCATGCTCTTCCGACATGCGGTGAGCGCTACTACAGAGACGCTTTCCGACAA 1418
QY 67 ValGlyAspValLeuArgAspThrLeuSerGlnGluGluIleAspAspPhe---ArgLeu 85
Db 1417 GTGGGGCCAACTCTTCCGAAGAGCTCGCGGTGAACCCGAAACACCGGTATCTCCAGCGCTTA 1358
QY 86 -----GluValAlaLeu-----ProProGlnLysThrAsp 95
Db 1357 CGACAGATCAAGTCGCTCCCGCTCCAGCAGAAAGATCTCTCCACAGCATCCAGCA 1298
QY 96 GlyLeu-----GlySerGlyAspGlnLysIleGluValSerGlnPro 111
Db 1297 GGTGTACAGCACCGCCCGGAAATGGGATGGTACTCCGTGAAAGCATCAACCAACT 1238
QY 112 CysProGlySerLysLysGlnLeuLysPheGlnGluLeuGlnCysAspValSerMetGlu 131
Db 1237 GCACATTCGAGGAGCGTCACTCGTCAGCGCTCGATCCGCGCATGTATCCGCAACTCCGG 1178
QY 132 GluAspSerArgGlnGluTrpThrPheThrLeuTyrAspPheAspAsnGlyLysVal 151
Db 1177 CCAAGATGTGGGCAAGGACGCGCAAGCAAGAAAGACACCAAGCGGTATATGCCGGAAGCAC 1118
QY 152 ThrArgGlnAspIleThrSerLeuLeuHisThrIleTyrGlnValValAspSerSerVal 171
Db 1117 CTACGCGCG----- 1109
QY 172 AsnHisSerProThrSerSerLysMetLeuArgValLysLeuThrValAlaProAsp--- 190
Db 1108 ---CATCTACAGAGATGATCACTTCTGACAAACCAAGCGCCCTTCAGACCGGACAC 1052
QY 191 ---GlySerGlnSerLysArgSerValLeuValAsnGlnIleAspLeuGlnSerAlaArg 209
Db 1051 CATGGGAGCGGTGCGCAACGTCGCGCT-----GATGGCGCGCA 1013
QY 210 ProArgAlaGluThrLysProThrGluAspLeuArgSerTrpGlu----- 224
Db 1012 GCGCGAGAAATACGCGTCCACGACGACCAAGCTTCCGAATATGACGCGCACCGCACCATGCG 953
QY 225 -----LysLysGlnArgAlaProLeuArgPheGlnGluAsp 236
Db 952 CGTGTCTCTGGCCGAGCGAGCGCTGCTGATGACAGCAGCAGTGGAGACCGCGCATCTG 893
QY 237 SerArgLeuGlnGlnSerGlyCysTyrHisHisCysValAspGluAsnIleGluArgArg 256
Db 892 GCGCGCGCTGCGCAACCAAGG-----CGCAC 866
QY 257 AsnHisTyrLeuAspLeuAlaGlyIleGluAsnTyrThrSer----- 270
Db 865 GATCCGCGAGTCTGACAGCTGCGCGCTCACCCGCGCAGTCCAGTCCAGTCCCGCGGAT 806
QY 271 ---GlnPheGlyProGlySer-----ProSerValAlaGlnLysSerGlnLeuPro 286
Db 805 CTCTTGGCTGGACCCGAGCGCGCCGACGACCGGAACTCGGAGAAAGGTCCAGCTGTA 746
QY 287 ProArgThrSerAspProThrArgSerArgSerHisGluProGluAlaIleHisIlePro 306
Db 745 CCTCAAGACCAAGACCTGACCGGTCTCGA---CATAGACATCATGGGCTCAACAGAGGC 689
QY 307 HisArgLysProGlnGlyVal---AspProAlaSerPheHisPheLeu----- 321
Db 688 CATTCGCTCAGCATGAAGCGCTGATCCGCGCAAGACACCATCTCGGTGACCGGCA 629
QY 322 -----AspThrProIleAlaLys-----ValSerGlnLys---Gln 332
Db 628 CGTCTCGCGGACCTACTGACCGACCTGTCTCCGATATGAACTGGGCACTCGGCGCA 569
QY 333 GlnArgLeuArgGlyThrGlnAspGlySerLysHisPheValArgSerProLysAlaGln 352
Db 568 GATGCTCTCATGTTCCGCTGATGGCGGCGCGGACGATTAAGAAACCGGCGCGCGG 509

```

```

QY 353 GlyLysSerValGlyValGlyHisValAlaArg-----GlyAlaArgAsnLysPro 369
Db 508 -----CTGGCACCACCAAGCAGTCCAGCAACTGGTGAAGAGAACTACT 464
QY 370 ProLeuGlyProAlaIleProAlaValSerProSerAlaHisIleValAlaSerProAla 389
Db 463 GCGCTGGGA-----CTCCCTGGCGGAGTCTCTGGC----- 434
QY 390 LeuLeuProSerLeuAlaProLeuGlnHis----- 399
Db 433 -----CTGGCGCTGTCTCTGGAGAAACCGGATCAAGACCGCAACCGCAAGGCCAA 380
QY 400 -----LysLysHisLysHisArgAlaLysGlnSerGln----- 410
Db 379 GCTGCTCGCAAGCGCTGACCAAGCCAGCCAGCAGCTGTGACAAACAAAGTCCGCC 320
QY 411 -----GlnGlyCysArgGlyLeuGln-----AlaProLeuAlaSerGly----- 423
Db 319 GTGGCGAAGGTGGCGCAATCGACAAACCGCGCAGCCACTTCTACTGGCGATGTACTG 260
QY 424 GlyProValLeuGlyArg 429
Db 259 GCGCCAGCGCTGCGCGC 242

RESULT 10
US-08-999-689A-2
; Sequence 2, Application US/0899689A
; Patent No. 654615
; GENERAL INFORMATION:
; APPLICANT: ULARICH, AXEL
; APPLICANT: KHARITONENKOV, ALEXEI
; APPLICANT: CHEN, ZHENGTUN
; TITLE OF INVENTION: SIRP PROTEINS AND USES THEREOF
; FILE REFERENCE: 038602/0548
; CURRENT APPLICATION NUMBER: US/08/999,689A
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/030,964
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2433
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-08-999-689A-2

Alignment Scores:
Pred. No.: 0.00224 Length: 2433
Score: 152.00 Matches: 140
Percent Similarity: 33.04% Conserved: 51
Best Local Similarity: 24.22% Mismatches: 189
Query Match: 6.08% Indels: 198
DB: 4 Gaps: 31

US-09-993-966-7 (1-470) x US-08-999-689A-2 (1-2433)
QY 3 LysLeuHisSerLysProAlaValAlaValCysLysArgArgGlnSerProGlnGlyAspSer 22
Db 507 GACCTTCACCTGGAGATGCCAGCGCTTCTCACCAGACAGATCACTACCTGA 557
QY 23 PheAlaValSerAlaIleAlaTrpAlaArgLysGlyIleGlnGluTrp 37
Db 558 -----ATGTTCAAAAATGGGAATAGCTTCAGA 587
QY 38 IleGlyArgGlnArgCysPro----- 44
Db 588 CTTCAGACCAAGCTGAGACCCCTAGAGAGACCGTCTCTACAGCATCCAGCACAGC 647
QY 45 GlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGlyArgSerThrArg 64
Db 648 CAAGGTGTGTCTACCCGCGAGGACGTTCACTCTCAAGTATCTGGAGGTGCGCACG 707
QY 65 GlnLeuVal-GlyAspValLeuArgAspThrLeuSerGlnGluGlnAspAspPhe 84

```

```

Db      708  CACTTGGACGAGGGAGACCTCTTGCGGACCTGCCAACTGTCGAG----- 753
Qy      84  gIeuGluValAlaLeuProGluIlyThrAerGlyLeuGlySerGlyAerGluIly 104
Db      754  ----ACCATCGAGTTCACCC----- 773
Qy      104  sMeGluArgValSerGluProCysProGlySerGlyLeuGluLeuIlyPheGluLeu 124
Db      774  CTTGGAGGTACTCAACAGCCCGCGAGGAGGAGAACAG-----GGAATGT 821
Qy      124  uGlnCyAerPValSerMetGluLeuSerAerGlnGluTrpThrPheThrLeuTyra 144
Db      822  CACTCGCAGGTGAGGAAGTTCTACCCCGAGAGCTACAGCTACGCTGGTTG----- 873
Qy      144  pHeAerAerAerGlyIlyValThrArgGluAerPheThrSerLeuLeuIlyThrIle 164
Db      874  ----GAGAAATGAAACGTGTCCTCGGACAGAAACGGCTCA-----ACCGTTAC 917
Qy      164  rGluValAlaAerSerSerValAerHisSerPro-----ThrSerLeu 179
Db      918  AGAGAACAGAGATGTACTCAACTGATGAGCTGCTCGTGAATGATATCTGCCCA 977
Qy      179  sMeGluArgValIlyLeuThr-----ValAlaProAerGlySerGlnSerIlyAer 197
Db      978  CAGGATGATGTGAACTCACTGCTGAGCTGAGCATGACGGGACGACGCTCAGCAA 1037
Qy      197  erValLeuValAerGlnAlaAerLeuGlnSerAlaAerProAlaGluThrIlyPro 216
Db      1038  AAGCCATGACTGAAAGGTCTCAG-----CCACCCGAAAGACAGGCTCAATACCGCGC 1094
Qy      217  ----ThrGluAerLeuAerSer----- 222
Db      1095  TGAGAACACTGATCTAATGAACGGAACATCTATTGTGTGTGTGTGTGTGTGTGTGT 1154
Qy      223  ----TTP-----GluIlyLeuGln 228
Db      1155  GCTGTGTCCTACTGATGCGGCGCTCTACTGCTGCCAATCAGACAGAAAGAACCCCA 1214
Qy      228  rGluAerProLeuAerPheGlnGly-----AerSerAerGluLeuGlnSerGlyCys 244
Db      1215  GGGGTCTCACTTCTTCAACAGGTTGATGAGCCGAGAGAGAGAGAGAGAAATACACA 1274
Qy      245  ----TyrHisHisCysValAerPgluAerHisLeuArgAerAerHisIlyLeu 260
Db      1275  GGCACAAATGATATCATATGACAGCTGAACCTGCCAA----- 1316
Qy      261  AsPleuAlaGlyIleGluAerTrpThrSerGlnPheGlyProGlySerPro--SerVal 279
Db      1317  ----GGGGAAGAGCTGCTGCCAGCTGCCGAGGCCAACAACACACAGAGTA 1367
Qy      280  AlAGlnIlySerGluLeuPro-----ProAerGlnSerAerPro 292
Db      1368  TGGCAGGATTCAGACAGCCGCGAGCCGCTGAGAGACACCTCATGCTGTAAGCT 1427
Qy      293  ThrArgSerAerSerHisGluProGluAlaIleHisIleProHisAerGlyProGlnGly 312
Db      1428  ----GACATGCTCA-----CCTCAACCGGAGCCCAAGCA 1460
Qy      313  ValAerProAlaSerPheHisPheLeuAerThrProIleAlaIlyValSerGluLeuGln 332
Db      1461  GCCGCGCCCAAGCC-----TGAGCCGTCTTCTCAGAGTACGC 1499
Qy      333  GluAerG-----LeuAerGlyThr--GlnAerGlySerIlyHisPhe--Val 346
Db      1500  CAGGTCACAGTCCCGAGAGAAATGAATGGACCGGTGTTGCTTAGACACCATCTCTAC 1559
Qy      346  IArgSerProIlyValAGlnGlyIlySerVal-----Gln 357
Db      1560  GCGCTTCTTCTTCCACAGGAGCCCGGTATATGACACACCAACCAAGTCCCGAGAG 1619
Qy      357  yValGlyHisVal-----AlaArgGlyAlaAerAerIlyProIle 371

```

```

Db      1620  GCTGGGCGGTGACAGCTCTGGAGCCAGGAGCCAGGCTGCTCTTCTCCCAACC-- 1677
Qy      371  uGluProAlaIleProAlaValSerProSerAlaHisIleValAlaAerProAlaLeu-L 391
Db      1678  ----CTCCTTGGCTCTCCACACTTCC 1700
Qy      391  euProSerLeuAlaProLeuGlyHisIlyValHisIlyHis-----ArgA 406
Db      1701  TGGGACGCCACGCGCCCTC-----CCCCAATGTCACACACCTGAGGCTGACGTT 1754
Qy      406  IAlaSerGluSerGln-----GlnGlyCysAerGlyLeuGlnAlaP 419
Db      1755  GCCAAACAGCCCGGAGAACCAACCTGGAGAACTGCGAGAACTGCGGGGTCCAGAACT 1814
Qy      419  roLeuAerSerGlyGlyProValLeuGlyAerGluHisLeuAerGlu--LeuProAlaLeu 438
Db      1815  CTTGTGCTCTCCCTCCATCACTACATGTGGTTTGAAGACCTGACAGTCCCTCCGAGTCT 1874
Qy      439  ValValTyGlu-----SerGlnAlaGlyGlnProValGlnAerHis 452
Db      1875  CCGAGCTGATCTTCCAGGCTGGGAGAGAAATCCCACTCCCTGACTTCCACAC 1934
Qy      453  GluHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 466
Db      1935  CTCACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1976

RESULT 11
US-09-944-807-3
? Sequence 3, Application US/09944807
? Patent No. 6773895
? GENERAL INFORMATION:
? APPLICANT: Boehringer Ingelheim Pharma KG
? TITLE OF INVENTION: Method for identifying substances which positively
? TITLE OF INVENTION: influence inflammatory conditions of chronic
? TITLE OF INVENTION: inflammatory airway diseases
? FILE REFERENCE: 082.00n
? CURRENT APPLICATION NUMBER: US/09/944,807
? PRIOR FILING DATE: 2001-08-31
? PRIOR APPLICATION NUMBER: UK 0021484.1
? NUMBER OF SEQ ID NOS: 24
? SOFTWARE: Patent Ver. 2.1
? SEQ ID NO 3
? LENGTH: 2433
? TYPE: DNA
? ORGANISM: Homo sapiens
US-09-944-807-3

Alignment Scores:
Pred. No.: 0.00224 Length: 2433
Score: 152.00 Matches: 140
Percent Similarity: 33.04% Conservative: 51
Best Local Similarity: 24.22% Mismatches: 189
Query Match: 6.08% Indels: 198
DB: 4 Gaps: 31

US-09-993-966-7 (1-470) x US-09-944-807-3 (1-2433)
Qy      3  LysLeuHisSerIlyProAlaAlaValCylAerAerGluSerProGluIlyAerSer 22
Db      507  GAGCTTCACTGAGAGTCCACGCGCTTCAACCGAGACATCACCTGAA----- 557
Qy      23  PheAlaValSerAlaAlaTrpAlaAerGlyIlyLeuGluIlyTrp----- 37
Db      558  ----ATGTTCAAAAATGGAAATGAGCTTCAGA 587
Qy      38  IleGlyAerGlnAerCysPro----- 44
Db      588  CTTCAACCAACAGTGAACCCCGTAGAGAGAGGTCTCTACAGATCCACAGACAGC 647
Qy      45  GlyGlyValSerGlyProAerGlnLeuAerGluAlaGlyThrIleGlyAerSerThrArg 64
Db      648  CAAGGTGTGTGTCGACCGGAGAGAGCTTCACTCAAGTCATCTCAGAGTGGCCACGT 707

```

```

QY      65  GluLeuVal-GlyAspValLeuArgPheThrLeuSerGluGluGluLeuAspPheAsp  84
DB      708  CACCTTGACGAGGAGACCTCTTCTGAGACCTGCCACCTTGTCTAG-----  753
QY      84  GluGluValAlaLeuProGluIuylsThrAspGlyLeuGlySerGlyAspGlyIuyls  104
DB      754  ----ACCATCCGAGTTCACCC-----AC  773
QY      104  sMetGluArgValSerGluProCysProGlySerIleGluLeuIuylsPheGluGluLe  124
DB      774  CTGGAGAGTACTCAACAGACCCCGAGGAGGAGAGAACAG-----GTGAATGT  821
QY      124  uGlnCysAspValSerMetGluGluIuylsAspSerArgGlnGluIuylsThrPheThrIuyls  144
DB      822  CACCTCCGAGGTAGAGAACTTCAACCCCAAGACTACAGTGAAGCTGAGTGG-----  873
QY      144  pPheAspAsnAsnGlyIuylsValThrArgGluAspIleThrSerLeuIuylsThrIleTy  164
DB      874  ----GAGAAATGAGAAAGCTGTCGCCGACAGAAACGGCTCA-----ACGTTAC  917
QY      164  rGluValAlaAspSerSerValAsnHisSerPro-----ThrSerSerTy  179
DB      918  AGAGAACAGAGATGATGACTACAACTGATGAGCTGCTCCTGTAATGATGCTGCCA  977
QY      179  sMetLeuArgValIuylsLeuThr-----ValAlaProAsp-GlySerGlnSerIuyls  197
DB      978  CAGGAGTGAATGATGAGCTCACTGCCAGTGAAGCATGAGGAGGAGGAGGAGCTCAGCAA  1037
QY      197  eValLeuValAsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIuylsPro-  216
DB      1038  AACCCATGACCTGAAGGTCTCAG---CCACCCGAAAGAGAGGAGGCTCAATACCGCGC  1094
QY      217  ---ThrGluAspLeuArgSer-----  222
DB      1095  TGAGAACACTGATCTAATGACAGCAACATCTATATTGTGTGTGTGTGTGTGTGTGTGTGT  1154
QY      223  -----TTP-----  228
DB      1155  GCTGTGGCCCTTACTGATGGCGGCGCTTACTCTGTCGATCAGACAGAAAGAAAGCCCA  1214
QY      228  rGluAlaProLeuArgPheGlnGly-----AspSerArgLeu-GluGlnSerGlyCys  244
DB      1215  GAGCTCACTTCTTCTTACAGAGGTTGCTGATGAGCCCGAGAGAAAGTCCAGAAATACACA  1274
QY      245  -----TyriHisCysValAspGluAsnIleGluIuylsArgAsnHisTyLeu  260
DB      1275  GAGACAAATGATATCATATGACAGACCTGAACTGCCCA-----  1316
QY      261  AspLeuAlaGlyIleGluAsnTyriHisSerGlnPheGlyProGlySerPro--SerVal  279
DB      1317  -----GGGAGAGAGAGCTGCTCCCGAGGCTGCGAGCCCAACCAACGAGGAGTA  1367
QY      280  AlaGlnIuylsSerGluLeuPro-----ProArgThrSerAsnPro  292
DB      1368  TGCACACATTCAACAGACCCCGAGCGCTGAGAGACACCTTCACTATGCTGACCT  1427
QY      293  ThrArgSerArgSerHisGluProGluAlaIleHisIleProHisArgIuylsProGlnGly  312
DB      1428  -----GACATGTGTCCA-----CTCAACCGGAGCCCGCAAGCA  1460
QY      313  ValAspProAlaSerPheHisPheLeuAspThrProIleAlaIuylsValSerGluLeuGln  332
DB      1461  GCGCGGCCCGCCAGCC-----TGACCCGCTCTTCTTCAAGAGTACG  1499
QY      333  GluArg-----LeuArgGlyThr---GlnAspGlySerIuylsHisPhe-Val  346
DB      1500  CACGCTCCAGGTCCGAGAGAGATGAGACCGTGTGCTTGTCTTACACCACTTCTTAC  1559
QY      346  IArgSerProIuylsAlaGlnIuylsSerVal-----Glu  357
DB      1560  GCGCTTCTTGTGCCACAGGAGCGCGCTGATGAGACAGCAACCCAGTTCGCGAGG  1619

```

```

QY      357  ValGlyHisVal-----AlaArgGlyAlaArgAsnIuylsProProLe  371
DB      1620  GCTGGGCGGCTGACAGGCTTGGGAGCCAGGAGGAGGAGGCTTCTTCTTCCCAACC-  1677
QY      371  uGlyProAlaIleProAlaValSerProSerAlaHisLeuAlaIaAspProAlaLeu-L  391
DB      1678  -----CTCTTGAGCTCTCCAGACATTTCC  1700
QY      391  euProSerLeuAlaProLeuGlyHisIuylsHisIuyls-----ArgA  406
DB      1701  TGGGACACCAACGCGCCCTC-----CCCAACATTGCCACACACTGAGAGGTGAGCTT  1754
QY      406  IalysGluSerGln-----GlnGlyCysArgGlyLeuGlnIaIaP  419
DB      1755  GCCAAACACCCAGGAGAACCAACCTGGGAAGTGGCCGAACTGCTGGGCTCCAGAACT  1814
QY      419  rGluAlaSerGlyGlyProValLeuGlyIuylsArgGluIuylsLeuProAlaLeu  438
DB      1815  CTGTGCTCTCCCTCCATCACCACATGTGGTTTGAAGACCTCCAGCTGCTCCCGATGCT  1874
QY      439  ValValTyrglu-----SerGlnAlaGlyGlnProValGlnArgHis  452
DB      1875  CCGAAGCTGATCTTCCAGGAGTGGGAGAGAAATCCACCTGCCCTGACCTCCACAC  1934
QY      453  GluHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis  466
DB      1935  CTCACACACACACACACACACACACACACACACACACACACACACACACACACAC  1976

RESULT 12
US-09-103-840A-2
: Sequence 2, Application US/09103840A
: Patent No. 6294328
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: TITLE OF INVENTION: TUBERCULOSIS
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103,840A
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 4403765
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: FEATURE:
: OTHER INFORMATION: CDC 1551
: OTHER INFORMATION: "n" bases at various positions throughout the sequence
: OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 296 Length: 4403765
Score: 149.50 Matches: 133
Percent Similarity: 29.76% Conservative: 67
Best Local Similarity: 19.79% Mismatches: 200
Query Match: 5.98% Indels: 272
DB: 3 Gaps: 30

US-09-993-966-7 (1-470) x US-09-103-840A-2 (1-4403765)
QY      9  AlAlaValAlaCysAlaArgArgGluSerProGluGlyAspSerPheAlaValSerAlaAla  28
DB      3034730  GCGCGCATACCGCAAGCGGCTCTCGACCA-----  3034759
QY      29  TrpAlaArgIuylsGlyIleGluGluIuylsArgGlnArg-----CysProGlyGly  46
DB      3034760  ---GTGCTGGCGGCTCCGAGTGAAGTGGCAACCGGCTGACAGTGAACCGGAGGT  3034816
QY      47  ValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGly-----  60

```

```

Db 3034817 CTGGGGAACCAAGACCCGCTGCTGACAGTACGGCGGCTGCCAGGTAAGTCTGCT 3034876
QY 60 -----
Db 3034877 CCGCGCGGCGCATCATGCTCAGCTCCAGCACCGCGCTCCAGAGCGGCTATTACTGCG 3034936
QY 61 ArgSerThrArgLysLeuVal-GlyAspValLeuArgAspThrLeuSerGlu----- 78
Db 3034937 CGTAGCCGACGCGCGCTGACGCGGTACCCCTCGCGCAGACCATTCGAGCAGAGTGC 3034996
QY 78 -----
Db 3034997 GTACCTTTCAACAGGCGCTGATCAACATCAGTGGTGCCAGCGGCGCAACCGCTGCT 3035056
QY 79 -----
Db 3035057 CGAGAAATCGTGTCTGACAGTCCAAACCGGACGATTAACCGT-----GTCCACCC 3035104
QY 91 OGILyysThrAsp-----GlyLeuGlySer-----GlyAspGlyLysLysSme 105
Db 3035105 GCGGCGACCGATGCTGCGCGCGAGCGGCAAAATGCTGCCGCTGATTCGACCACT 3035164
QY 105 tGluArg-----ValSerGluProCysAspProGlySerLysLysGlu 118
Db 3035165 CGAGCGCGCCGACCGTGCAGCGCGCATCTGTGGCGCAGAGATTGCCGCGCTGCGACCGGGT 3035224
QY 118 n-----LeuLysPheGluGluGluGlnCysAspVal-SerMet----- 130
Db 3035225 CTCGCGCGGCTAACTCGGCGCTGACGCGCATCCCGACCTCGGCGCAGCGCGCTCCC 3035284
QY 131 -----
Db 3035285 ATCTCGCGCTACTGAAAGATCGGTTGCGGAAACGACAGTCTGACGAGGATTGGA 3035344
QY 139 hrPheThrLeuTyArgPheAspAsnLysValLysValLysGluAspLysLysSerL 159
Db 3035345 CATACACATCGACGCGCGCCGACCGACCGCGCTCCCGGCTGCGATCCCTCGA 3035404
QY 159 eu----- 159
Db 3035405 TGTCGCGCGCGCGCGCGCTGTAGCGCGCACGCTGCGGTTTCGGTATCATCCAGA 3035464
QY 160 -----
Db 3035465 CATCGAGTTAATGATGCGGAGATCCACGCGCGCTGACGCGGAGCTTCCGCTCCCGA 3035524
QY 168 spSerSerValAsnHisSerProThrSerLysMetLeuArgValLysLeuThrValA 188
Db 3035525 TGTCATGCGCGCATACAGCTGATGCGCTGCGGTTCAAGATCTCCACGCTCACTCG 3035584
QY 188 laProAspGlySerGlnSerLysArgSer-----V 198
Db 3035585 CGCGGACCGCGCGCGCATCGAGCGCAACTGGGACTTCCGCGCGCGCTGCTCGGA 3035644
QY 198 allLeuValAsnGlnAlaAspLeuInserAlaArgProArgAlaGluThr----- 214
Db 3035645 TAAATGCCAAC-----GTTTACGCGCTGCCAGACCGCGCAATATACCCACGCGGTG 3035698
QY 215 -----
Db 3035699 TGGCGCTCTCGTGAAGACTCTTGGCGGATCGCGGCTCGGCTCGGCGACGCGGCGCAG 3035758
QY 217 hrGluAspLeu-----A 221
Db 3035759 CACCTGAATGTCACCGCGCCACACCTGTGCGGAGCGCGGCTCAAGACGCGAGGTC 3035818
QY 221 rgSerTrpGluLysLysGlnArgAlaProLeu-----ArgPheGlnLysAspSerA 238
Db 3035819 GCGGCTGACCAACGCGCTGCGAAGCGCGCTGATACCGGCGCGCTGAGGCTATAGCC 3035878
QY 238 rgLeuGluGln-----SerGlyCysTrpTyrHisLysCysValAspGluAsnLysGluArgA 256

```

```

Db 3035879 GCGGAGACGCGGCGCTGAGGCTGT-----TCAGCGCGTC 3035914
QY 256 rgAsnHisTrpTyrLeuAspLeuAlaGlyLysGluAsnTyrThrSerGlnPheGlyProGlyS 276
Db 3035915 GCGGACGACTAGTACC-----CGATGATCGGCTCCGCTC 3035953
QY 276 erProSerValAlaGlnLysSerGluLeuProProArgThrSerAsnProThrArgSerA 296
Db 3035954 GATTCGCGCTCGCAGTTGCTGCGCAGTGCACCGCGGCGATCGGAC-----CGTGTCT 3036007
QY 296 rgSerHisGluProGluAlaLeuHisLysProHisArgLysPro----- 310
Db 303608 GCTGCATAGACGTGACGCTGCTCCCTGCTGAGCGCGGCGCTGCGCCGTCACCAAGC 3036067
QY 311 -----
Db 303608 GGTAACTAGCGCACACAGGGGTAAATTCACCGGACGCGCACTCCGTACCGCTCTCGG 3036127
QY 326 laLysValSerGluLeuGlnArgLysGluArgLysThrGlnAsp-----G 341
Db 3036128 GCGGA-----GCCGACCTCGACGTCGCGCGCGAAGCCCGCAAGAGTCCGACGTCGTTGGG 3036184
QY 341 LysSerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValGlyHisV 361
Db 3036185 GCGCACGACGTCGTCGCGCGCGCGCTCCGACGACGATCCAGCTTT----- 3036231
QY 361 alAlaArgGlyAlaArgAsnLysProLeuGlyProAlaLeuProAlaValSerProS 381
Db 303632 -----TCGGACACAGAGGCGCGCGGATCACCG-----CGGCCC 3036268
QY 381 erAlaHisValAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLysL 401
Db 3036329 CAAATCGCGCAGCTCGGACGCGCGGCTCGGCGCATTCGGAACCAACCGCGCGGA 3036328
QY 401 yHisLysHisArgAlaLysGluSerGlnGlnGlyCysValGlyLeuGlnAlaProLeuA 421
Db 3036329 CGGACGATCCGATGCGCTCAACACACAAAGCTTAACGCTTG----- 3036375
QY 421 laSerGlyGly-----Prov 426
Db 3036376 -GCGGCGGATTCGCGCTGTCGAGACACGCGTGCACAAAGACCGTCTGAGAGCGCGA 3036433
QY 426 allLeuGlyArgGluLysLeuArgLysLeuProAlaLeuValLysGluSerLysLag 446
Db 3036434 CACTGTCCCGCGCTCATGCGCGAG-----ACGTCAACGAAACCGCACTG 3036481
QY 446 ly-----GlnProValGlnArgHisGlu-HisHisHis----- 457
Db 3036482 GCCCGCGCGCCCAATTCGCGCGGAAACCGTCAATCATGCGACCTGTCACGGCGCA 3036541
QY 458 -----HisGluHisHisHisTyr 464
Db 3036542 GCGTACGTCGCGACCAACACACACTAC 3036571

```

RESULT 13
 US-09-103-840A-1
 Sequence 1, Application US/09103840A
 Patent No. 6294328
 GENERAL INFORMATION:
 APPLICANT: FLEISCHMAN, Robert D.
 APPLICANT: WHITE, Owen R.
 APPLICANT: FRASER, Claire M.
 APPLICANT: VENTER, John C.
 TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 FILE REFERENCE: 24366-20007.00
 CURRENT APPLICATION NUMBER: US/09/103,840A
 NUMBER OF SEQ ID NOS: 2
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 1
 LENGTH: 4411529
 TYPE: DNA

```

/ ORGANISM: Mycobacterium tuberculosis
/ OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.: 297 Length: 4411529
Score: 149.50 Matches: 133
Percent Similarity: 29.76% Conservative: 67
Best Local Similarity: 19.79% Mismatches: 200
Query Match: 5.98% Indels: 272
DB: Gaps: 30

US-09-993-966-7 (1-470) x US-09-103-840A-1 (1-4411529)

QY 9 AAlAAlValCyblybAArgluserProgluYlApsSerPheAlaValSerAlaA1a 28
DB 3039833 GCGGCGCATACCGCACCGGCGCTCGTCAGCA----- 3039862
QY 29 TrpAlAArglybglYllegluTrpIleglYArglnArg-----CySproglYglY 46
DB 3039863 ----GTGTCGGCGCGCTCGGTAGCTGGCCACACCGGCGTGAAGCCAGTCACCCGGTGT 3039919
QY 47 ValSerGlyProArglnLeuArglnAglYThrIleglY----- 60
DB 3039920 CTCGGCGAAACAGACCGCTGCCGTGCAAGTACGGCGGGTCCAGGTAGCTGTCT 3039979
QY 60 ----- 60
DB 3039980 CCGCGCGCGCGCATATGTCTCAGCTTCACACCGCGCTCCAGACGGCTATTACTGCG 3040039
QY 61 ArgSerThrArglnLeuVal-GlyApsValLeuArgApsThrLeuSerGluGlu----- 78
DB 3040040 CGTACCGACCGCGCGGTGACCGGTGACCCCTCGCCAGACCATTCGCGAGCAGAGTGC 3040099
QY 78 ----- 78
DB 3040100 GTACTCTTCAACACGAGCCCTGATCAACATCAAGTGTGTCGACGCGGCCAACCCGTCT 3040159
QY 79 -----GluGluApsApsPheArgLeuGluValAlaLeuProP 91
DB 3040160 CGAGATCTGTGTCTGACAGTCCACCCGACGATACCGT-----GTCCACCC 3040207
QY 91 ogluYbThrAps-----GlyLeuGlySer-----GlyApsGluYblybSme 105
DB 3040208 GCGGCGACCGATGCTGTCGCGCGCGGCAAGTGTGCGGTGATTCAGACCACT 3040267
QY 105 rgluArg-----ValSerGluProCyApsProGlySerlyblybgl 118
DB 3040268 CAGGCGCGCGACCGTGGCGCGGGCATCTGTGGCGAGATTGCGCGGTGCAAGCCGGT 3040327
QY 118 n-----LeuYbPheGluGluLeuGlnCyApsVal-SerMet----- 130
DB 3040328 CTCGCGCGGCTAACTCGCGGTGACGCCGATCCACCCGACCTCGCGCAAGCGCGTCCC 3040387
QY 131 -----GluGluApsSerArglnGluTrp 139
DB 3040388 ATCTCGCGGTACTGAGAGATCGGTGGGAAACACACCGTGTGAGACGAGGATTGGA 3040447
QY 139 hrPheThrLeuYbThrApsPheApsApsnglyblybValThrArgGluApsIleThrSerly 159
DB 3040448 CATACAGCATCGACGCGGCCACACAGACCGGCAACCGCTCCCGGGGTGCGATGCTCTCA 3040507
QY 159 eu----- 159
DB 3040508 TGTCCCGCGCGCGCGCGCTGTAGCGGCGCACGGTCCGGTTTCGATGATCCAGGA 3040567
QY 160 -----LeuYbThrIleYbGluValVala 168
DB 3040568 CATCGAGTTGATGATGCGGAGATCCACGGCGCTCGCTGACGCGGACGCTTCGCCCA 3040627
QY 168 psSerSerValApsnHisSerProThrSerSerlybMetLeuArgValYbLeuThrVala 188
DB 3040628 TGTTCATGCGCGGATACAGCTGATGCGGTGCGGTTCACAGATCTCACGCTCACTCTGG 3040687

```

```

QY 188 lApsApsGlySerGlnSerlybArgSer-----V 198
DB 3040688 CGCGGACCGCGCGCGCATCGACCGCAACTGGACTTGGCGCGCCCGTCTGTCGA 3040747
QY 198 alLeuValApsnglnAlaApsLeuGlnSerlybApsProApsArgAlaGluThr----- 214
DB 3040748 TAATCGCAAC-----GTTCTACAGCTGCGACACCGCGAAATATCCACAGCCGTG 3040801
QY 215 -----LybProT 217
DB 3040802 TGGCGTCTCTCGGTAGAACTCTTGGCCGATCGCGGTCTGTGCTCGGCGACGCGCGCAG 3040861
QY 217 hrGluApsLeu-----A 221
DB 3040862 CACCTGAATATGCCACCGGCCAACACACTGTGGGAGCGCGGTCAAGACGGCGAGTTC 3040921
QY 221 rGSerTrpGluYblybGlnAArgAlaProLeu-----ArgPheGlnGlyApsSerA 238
DB 3040922 GCGGCTGGCCAAAGCGCTGTCTGAGAGCCCGTGCATACCGGCGCGGTGATATAGCC 3040981
QY 238 rGLeuGluGln-----SerGlyCybTybHisbCybValApsGluApsnIlegluArgA 256
DB 3040982 GCGGAGAGCGCGGGCGGTCAAGGTGT-----TCAAGCGGTCTC 3041017
QY 256 rGApnHisTybLeuApsLeuAglYllegluApsnTybThrSerGlnPheGlyProGlyS 276
DB 3041018 GCGGACGACTATGATCC-----CGATCGGATCGGCGCTCCCGTCTC 3041056
QY 276 exProSerValAglInlybSerGlnLeuProProApsThrApsApsProThrApsSerA 296
DB 3041057 GATGTGCGCTCGCATGTTCCCTGCGACGTGCGCGGCGCATTCGAGAC-----GTGCTC 3041110
QY 296 rGSerHisGluProGluAlaIleHisIleProHisArglybPro----- 310
DB 3041111 GCTGGCATTAAGTGAAGCTGTGCTGCTGCTCAGGCGCGGCGTCCGACCCGTAACCAAGC 3041170
QY 311 -----GlnGlyValApsProAlaSerPheHisPheLeuApsThrProIleA 326
DB 3041171 GGTAACTAGCGCACACAGGGGTAAATTCACCGAACGCGCAATCCGTCAACCGTCTGCGG 3041230
QY 326 lAlybValSerGluLeuGlnAArgLeuArglyblybThrGlnAps-----G 341
DB 3041231 CGGCA---GCCGACCTCGCACGTGCGCGCGGAAAGCCCGCAAGGTCCGACGTGCGTGG 3041287
QY 341 lySerlybHisPheValApsSerProlybApsIleGlnGlybSerValGlyValGlybHisv 361
DB 3041288 GCGCAGAGCGTCTCGCGCGCGCGGTCCGACAGCAATCAGCTTT----- 3041334
QY 361 alAlaArgGlyAlaArgApsnlybProProLeuGlyProAlaIleProAlaValSerProS 381
DB 3041335 -----TCGGCAACAAAGAGCGCGCGGATCACCG-----CGGCGCC 3041371
QY 381 erAlaHisIleuAlaAlaSerProAlaLeuLeuProSerIleuAlaProLeuGlybHislybL 401
DB 3041372 CAATCGGCGCAAGTTCGCGACAGCGCGGTCTCGGCAATTCGGAACCAACACCGCGCGGA 3041431
QY 401 ybHisIlybHisArgAlaYblybGlybSerGlnGlybCybArgGlybLeuGlnAlaProLeuA 421
DB 3041432 CGGAAAGATCCGATGCGGTCTCAACAAACAAAGCTAAAGCTTG----- 3041478
QY 421 lAserGlyglY-----ProV 426
DB 3041479 --GCGGCGGATTCGCGCTCTGTGAGCAACCGGTGCAAGAGACCGTGTGAGAGCGCGA 3041536
QY 426 alLeuGluYbArgGluHisIleAArglybLeuProAlaLeuValYblybTrpGluSerGlnAgl 446
DB 3041537 CACTGTCCGCGGTCTATGGCCAGAG-----ACGTCACAAAGACCGCAAGCTG 3041584
QY 446 ly-----GlnProValGlnAArgIleglu-HisbHisbHisv----- 457
DB 3041585 GCGCGGCGCGCCCAATTCGCGCGGGAACCGGTGATCATCGGACCTGTGACGGGCGCA 3041644

```

```

Db      3041645  GGCTACGGTCGCACACACCACCACTAC  3041674
RESULT 14
US-08-471-033-35
: Sequence 35, Application US/08471033
: Patent No. 5770696
: GENERAL INFORMATION:
: APPLICANT: Warren, Gregory W
: APPLICANT: Kozziel, Michael G
: APPLICANT: Mullins, Martha A
: APPLICANT: Nye, Gordon J
: APPLICANT: Carr, Brian
: APPLICANT: Desai, Nalini M
: APPLICANT: Kostichka, N. Kriсты
: APPLICANT: Duck, Nicholas B
: APPLICANT: Estruch, Juan J
: TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
: NUMBER OF SEQUENCES: 50
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CIBA-GEIGY Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30B
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/471,033
: FILING DATE:
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/314,594
: FILING DATE: 09-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/218,018
: FILING DATE: 23-MAR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/037,057
: FILING DATE: 25-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Pace, Gary M.
: REGISTRATION NUMBER: P-40,403
: REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8582
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 35:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2576 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "Synthetic DNA"
: HYPOTHEITICAL: NO
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 9..2564
: OTHER INFORMATION: /note= "Naire optimized sequence
: OTHER INFORMATION: encoding VPIA(a) with the Bacillus secretion signal removed
: OTHER INFORMATION: contained in pCIB5526"
US-08-471-033-35

Alignment Scores:
Pred. No.: 0.00728 Length: 2576
Score: 146.50 Matches: 101
Percent Similarity: 33.33% Conservative: 36

```

US-09-993-966-7 (1-470)	x	US-08-471-033-35 (1-2576)	Best Local Similarity: 24.57%	Query Match: 5.86%	Mismatches: 115
DB:	1	Indels: 160		Gaps: 21	
QY	190	AspGlySerGlnSerLysArgSerValLeuValaIleuAlaAspLeuGlnSer-----	207		
DB	56	GATGACCGCGAAGGGCCCTGCTGGCTACTTCAAGGGCAAGACCTTCAGCAACTCAG	115		
QY	208	--AlaArgProArgAlaGluThrLysProThrGluAspLeuArgSerTrpGluLysLys	226		
DB	116	CATGTGCGCCCGCAACGGGTGA CAGACCCCT-----GATCTCAGACAGACCGCCCAA	169		
QY	227	GlnArgAla-----ProLeuArg	232		
DB	170	CAAGCTGCTGGACAAGACGACGAGAGTACCGCATCCGCTGAATCGGCTGTATCCA	229		
QY	233	PheGlnGlyAspSerArgLeuGlu-----	240		
DB	230	GAGCAAGGAGACCGGCGGACTTCACTTCAACTGACGAGACGACGAGCCATCATCCA	289		
QY	240	-----	240		
DB	290	GATCAACGGCAAGATCATCAGACAAAGGCAAGAGACGATGTCACCTGGAGAA	349		
QY	241	--GlnSerGlyCybTrpThiHisCybValAspGluAsnIleGluArgArgAsnHisIyr	259		
DB	350	GGGCAAGCTGGTGCCCATCAAGATCGATACCGAGCGCACACCAAGTTCATCATCGACAG	409		
QY	260	LeuAspLeuAlaGlyIleGluAsnIyr-----	268		
DB	410	CAAGCACTTCAAGAGCTGAAGCTTTTCAAGATCGACGCCCAAGACCGCCACGAGGT	469		
QY	269	-----ThiSerGlnPheGly---ProGlySerProSer	278		
DB	470	GCAGAGAGACGAGCGCGCCACCCCGAGTTCAACAAAGAGAGACCGAGAGTCTTGCC	529		
QY	279	ValAlaGlnLysSerGluLeuProProArgThrSerAsnProThrArgSer---ArgSer	297		
DB	530	CAAGCCCGACAGATCAAA---CTGTTCACCCAGCAGATGAAGCCGAGATCGACAGGA	586		
QY	298	HisGluProGluAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSer	317		
DB	587	CACCGA-----CACCGACGGCAGACAGCATCCCGGACCTGTGGGA	625		
QY	318	PheHisPheLeuAspThrProIleAlaLysValSerGluLeuGlnIleArgLeuArgGly	337		
DB	626	GGAGAACGGCTACACCATCA-----GAACCGCATTCGCGTGAAGTGGGA	670		
QY	338	ThrGlnAspGly-SerLys-----HisPheValArgSerProLysValGlnGlyLys	354		
DB	671	CGACAGCTGGGTACCAAGGGCTACACCAAGTTCTGTAGCAACCCC---CTGGAGAGCCA	727		
QY	354	AspValGlyValGlnHisLeuAla-----AlaArgGlyAlaArgAsnLysPro-----	369		
DB	728	CACCGTGGCGCAACCCCTACACCGACTACGAGAGAGCGCGCGCGACCTGAGCTGACAA	787		
QY	370	-----ProLeuGlyProAlaIleProAlaIle-----	378		
DB	788	CGCCACGAGAGACTTCACACCCCTGTGTGGCGCCCTTCCCGACGCTGAACSTGACATGGA	847		
QY	379	-----SerProSerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLe	394		
DB	848	GAAGGTGATTCCTGAAGCCCAACGAGAACATGTGAGCAACAGCGT-GGA-----	892		
QY	394	uAlaProLeuGlnHisLysLysHisLysHisValArgAlaLysLysGlnSerGlnGlnGlyCybArg	414		
DB	893	-GAGCACTCGAGACCAACTGAGCTACACCAACACCGAGGGCGCGACGCTGAGAGCCG	951		
QY	414	GlyLeuGlnAlaProLeuAlaSerGlyLysProValLeuGlnLysArgGlnHisLeuArgGln	434		
DB	952	GCAT-----CGGTCCCAAGGAGATCAAGCTTGGCGGTGAG-----CTGCA	990		


```
QY 370 -----ProLeuGlyProAlaIleProAlaVal----- 378
Db 788 CGCCAGAGAGACCTTCAACCCCTGCTGGCCGCTTCCCAAGCTGAACTGAGCATGGA 847
QY 379 -----SerProSerAlaHisIleuAlaAlaSerProAlaLeuLeuProSerLe 394
Db 848 GAAGGTATCTCTGAGCCCAACGAGAACTGAGCAACAGCGT-GGA----- 892
QY 394 uAlaProLeuGlyHisIleuIleuHisIleuAlaIleuGlySerGlnGlnGlyCysAr 414
Db 893 -GAGCCACTCGAGCACCACCTGAGCTACACCAACACCGAGGCGCCAGCGTGGAGGCCG 951
QY 414 gGlyLeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlyArgGlyHisLeuArgG 434
Db 952 GCAT-----CGTCCCAAGGCGCATCAGCTTCGGCGTGA-----CGTGA 990
QY 434 uLeuProAlaLeu-----ValValTyrGlySerGlnAlaGlyGlnPr 448
Db 991 ACTACGACGACAGAGAACCGGTGCCAGAGAGTGGGCGACAGCACCGGCAACACGAGCC 1050
QY 448 oValGln----- 450
Db 1051 AGTTCAACACCGCGAGCGCGCTACCTGAACGCCAACGTGCGCTACACACGCTGGGCA 1110
QY 451 -----ArgHisGlyHisHisHis-----HisHisGly 459
Db 1111 CCGGCGCCATCTTAAGAGCTGAAGCCCAACCAAGCTTCGTGCTGAACACGACACATCG 1170
QY 459 uHisHisHisHisIleuTyrHisHisPheTyrGln 469
Db 1171 CCAACATCAACCGCAAGTCGAATTCACCGC 1201
```

Search completed: December 30, 2004, 09:00:02
Job time : 3940 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 30, 2004, 03:20:04 ; Search time 5644 Seconds
(without alignments)
3034.492 Million cell updates/sec

Title: US-09-993-966-7
Perfect score: 2499
Sequence: 1 MGLHKKPAVACVKRRSPG.....RHHHHHHHHHHHHFYQT 470

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xip
-Q/cgmr2.1/USPTO.spool.p/US09993966/runat.29122004.130758.9009/app.query.fasta.1.647
-DE=EST -OPMT=fastap -SUPPFX=rcf -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09993966.@CGM.1.1.6425.@runat.29122004.130758.9009 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEOUTERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_ges1:*
9: gb_ges2:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2489	99.6	2142	3	AF289584 Homo sapi
2	2179.5	87.2	2289	3	AK082367 Mus muscu
3	2094.5	83.8	2471	3	AK076164 Mus muscu
4	1852.5	74.1	1466	3	AK013241 Mus muscu
5	1545	61.8	885	9	AY412098 Homo sapi
6	1462	58.5	921	5	B0653673 AGENCOURT
7	1436	57.5	921	5	B0653673 AGENCOURT
8	1420	56.8	990	5	B0646566 AGENCOURT
9	1397	55.9	895	5	B0644360 AGENCOURT

10	1382	55.3	928	5	B0644956 AGENCOURT
11	1374	55.0	888	9	AY412100 Mus muscu
12	1283.5	51.4	1008	5	B0646371 AGENCOURT
13	1252	50.1	996	5	B0652087 AGENCOURT
14	1240	49.6	885	9	AY412099 Pan trogl
15	1216	48.7	966	5	B0649813 AGENCOURT
16	1195.5	47.8	915	6	CB182455 AGENCOURT
17	1186.5	47.5	1027	4	BG104777 AGENCOURT
18	1118	44.7	899	5	B0645507 AGENCOURT
19	1092	43.7	913	5	B0895164 AGENCOURT
20	1078.5	43.2	953	5	B0846054 AGENCOURT
21	1063	42.5	674	6	CB056689 AGENCOURT
22	1060	42.4	677	7	CF731367 UT-M-G20
23	1031	41.3	627	7	BI767278 AGENCOURT
24	1028	41.1	777	4	BG542261 AGENCOURT
25	1021	40.9	724	6	CA328380 UT-M-FYO-
26	1000.5	40.0	946	6	BY712252 AGENCOURT
27	978	39.1	839	7	CK795599 AGENCOURT
28	975	39.0	1059	5	B0070932 AGENCOURT
29	974	39.0	564	4	BM711145 AGENCOURT
30	918	36.7	653	2	BF540602 AGENCOURT
31	911	36.5	678	2	BF537335 AGENCOURT
32	897.5	35.9	825	6	CB961961 AGENCOURT
33	885	35.4	934	2	BF167269 AGENCOURT
34	867.5	34.7	2399	3	AK008814 Mus muscu
35	857	34.3	751	4	BG820139 AGENCOURT
36	855	34.2	622	6	CB215756 NISC np09
37	839.5	33.6	600	4	BI990034 B1054-43 M
38	830	33.2	540	5	BX513216 AGENCOURT
39	818.5	32.8	693	5	BM935129 UT-M-BH3-
40	811	32.5	670	2	BB650914 AGENCOURT
41	810	32.4	2802	3	AK081904 Mus muscu
42	808.5	32.4	684	7	CK450648 AGENCOURT
43	803	32.1	481	4	BI047069 MR3-FPN20
44	797.5	31.9	826	6	CB961668 AGENCOURT
45	793	31.7	522	4	BI903274 AGENCOURT

ALIGNMENTS

RESULT 1	AF289584	2142 bp	mRNA	linear	HTC 01-JAN-2002
LOCUS	AF289584				
DEFINITION	Homo sapiens clone pp7246 unknown mRNA.				
ACCESSION	AF289584				
VERSION	AF289584.1	GI:18027371			
KEYWORDS	HTC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 2142) Zhang,P.P., Zhou,X.M., Jiang,H.Q., Huang,Y., Qin,W.X., Zhao,X.T., Wan,D.F. and Gu,J.R.				
TITLE	Novel human cDNA clones with function of inhibiting cancer cell growth				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 2142) Zhang,P.P., Zhou,X.M., Jiang,H.Q., Huang,Y., Qin,W.X., Zhao,X.T., Wan,D.F. and Gu,J.R.				
AUTHORS	Man,D.F. and Gu,J.R.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-JUL-2000) National Laboratory For Oncogenes & Related Genes, Shanghai Cancer Institute, 25/In 2200 Xie-Tu Road, Shanghai 200032, P. R. China				
FEATURES	Location/Qualifiers				
source	1..2142				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="pp7246"				
	277..1689				
	/codon_start=1				
	/product="unknown"				

```

/protein_id="AAL55768.1"
/db_xref="gi:18027372"
/translation="MFKLHSPKPDVCKRRSPGSDSPAVSAMARKIEIKIGRQRC
GGVSPQLRLAGTIGRSTRELVDVLRITLSEEDFLREVALPEKIDGLSGGE
KMERVSEPCPSKKQLKFEELCDVMEBDSROEMFTVLVDNNKVRDITSL
HTTIEVDSVSNHPTSSKMLRVLTVAPOGSKRSLVNOADLOSPRAETKPE
DLSEWKEKORAPLRFQDGRLEQSGCHHCNDENIERNNYLDLAGIENYSGPDS
SVYAKSELPRTSNPTRRSHPEALHTRKQGVDPASFLDPTPIAKVSELOK
LRGTQDSKIFVSPKAGSKSVGVHVARGNRPGLPAIPAVSPAHIAAPLALP
SLAPLGHKKHHRKPSQDCRGLOAPLAGGVPVLRHLRELPAVVESQAGPVO
RHEHHHHHHHHYHFYQT"

```

ORIGIN

Alignment Scores:

```

Pred. No.:      5.6e-191      Length:      2142
Score:          2489.00      Matches:      468
Percent Similarity: 99.57%      Conservative: 0
Best Local Similarity: 99.57%      Mismatches: 2
Query Match:      99.60%      Indels:      0
DB:              3          Gaps:      0

```

US-09-993-966-7 (1-470) x AF289584 (1-2142)

```

QY      1  MecGIyLysLeuHisSerTysProAlaAlaValCysLysArgArgGluSerProGluGly 20
      277  ATGGGGAACCTTCACTCCAGCCCGGTGCACTGTGCAGAGCGAGAGAGCCCGGAAGCT 336
QY      21  AspSerPheAlaValSerAlaAlaTTPAlaArgLysGlyIleGluGluTTPleGlyArg 40
      337  GACACGCTTCGGCCGTGAGCGCTGCTGGGCTCGGAAGGCACTGAGAGTGGATCGGAGA 396
QY      41  GluArgCysProGluYglValSerGlyProArgGluLeuArgLeuAlaGlyThrIleGly 60
      397  CAGGCGTCCCGGGCGGGGTCTCGGAGACCCGAGACGCTGCGGTGGGGGACCAATGAGC 456
QY      61  ArgSerThrArgGluLeuValGlyAspValLeuArgSerThrLeuSerGluGluGlu 80
      457  CGAAGCACCCCGGAGCTGTGGGCGACGTGTGAAGACACGCTCAGAGGAGGAAGAGAG 516
QY      81  AspArgPheArgLeuGluValAlaLeuProProGluLysThrAspGlyLeuGlySerGly 100
      517  GACCACTTTCGGCTGGAAGTGGCCCTGCTCTGAGGAAGACTGACGGGCTGGGACAGCGA 576
QY      101  AspGluLysLysMetGluArgValSerGluProCysProGlySerLysLysGluLeuLys 120
      577  GATGAGAAAGAGATGAGAGAGTGAAGCACTCCAGGCTCCAGAGCACTGAGAG 636
QY      121  PheGluGluLeuGlnCysAspValSerMetGluLysAspSerArgGlnGluTTPThrPhe 140
      637  TTGTAAGAGCTCCAGTGGAGCTGTCCATGAGAGAGACAGCCGGCAGAGTGAACCTTC 696
QY      141  ThrLeuTyrAspPheAspAsnAsnGlyLysValThrArgGluAspIleThrSerLeuLeu 160
      697  ACCCTGATATGACTTTGACCAACAGCGCAAGGTCAACCGAGAGGACATCACAGACTTCGT 756
QY      161  HisThrIleTyrGluValAlaAspSerSerValAsnHisSerProThrSerSerLysMet 180
      757  CACACCACTATGAGGTGGTGAATCTCTGTCAACCACTCCCAACATCCAGCAAGAGTG 816
QY      181  LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysArgSerValLeuVal 200
      817  CTGCGGGTAAGCTCACCGTGGCCCCGATGGCAGCCAGAGCAAGAGAGCTCTTGTCTC 876
QY      201  AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeu 220
      877  AATCAGGCTGACCTGCAAGAGCGCAAGGCCCGGACAGAGCAACCACTGAGAGACTTG 936
QY      221  ArgSerThrGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
      937  CGGAGCTGGGAGAAAGACAGAGCCCGGCTCAGGTTCACGGGGTGAACGCGCTGGAG 996
QY      241  GluSerGlyCysTyrHisHisCysValAspGluAsnIleGluArgArgAsnHisTyrLeu 260

```

```

Db      997  CAGTCTGGCTGTACCAACCATGCTAGATGAGAACATCGAGAGAGAAACCACTTAA 1056
QY      261  AspLeuAlaGlyLysIleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValAla 280
Db      1057  GATCTGCCCGGATAGAAACCTACAGTCCCAATTTGGGCTGCTCCCTTCCTGGAGCC 1116
QY      281  GluLysSerGluLeuProProArgTyrSerAsnProThrArgSerArgSerHisGluPro 300
Db      1117  CAGAGTCAAGACTGCCCCCCCCGACCTCCATTCCTCACTGACTGCTCCATGAGCCG 1176
QY      301  GluAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPhe 320
Db      1177  GAAGCATTCACATCCACACCGAAGCCCAAGGCGTGAACCGGCTCTTCCACTTC 1236
QY      321  LeuAspThrProIleAlaLysValSerGluLeuGlnGluArgLeuArgGlyThrGlnAsp 340
Db      1237  CTTGACACCCCAATCGCAGAGGTCTCAGACTCCAGCACGGCTCCGGGGACCCAGAG 1296
QY      341  GlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValGlyHis 360
Db      1297  GGGAGCAGACACTTTGTGAGGTCCCCAGAGGCCCAAGGCAAGAGTGGGTGGGCGAC 1356
QY      361  ValAlaArgGlyValArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
Db      1357  GTGGCCAGAGAGGCAAGAAACAGCCCTCTGTGGACCCGCAATCCCTGCGGTGCCCC 1416
QY      381  SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
Db      1417  TCGGCCACCTGGCTGCGACCCGGCCCTCTCCCTCCCTACCCCTCCGCGGACACAG 1476
QY      401  LysHisLysHisArgAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
Db      1477  AAGCACAAGACCCGAGGCCAGAGAGAGCCAGAGGGTCCGGGGCTCGAGGACCACTG 1536
QY      421  AlaSerGlyGlyProValLeuGlyArgGluHisLeuArgGluLeuProAlaLeuValVal 440
Db      1537  GCTCAAGTGGCTGCTGTCTGGGGCGGAGCACTGTGGAGAGCTGCGCCGCTTGTTGG 1596
QY      441  TyrGluSerGlnAlaGlyGlnProValGlnArgHisGluHisHisHisHisHisHis 460
Db      1597  TATGAGGCCAGGCGCGGCGGCGGTCCAGAGCAATGAGCACCAACCAATGAACAT 1656
QY      461  HisHisHisTyrHisHisPheTyrGlnThr 470
Db      1657  CACCACATTAACCACTTTCACAGACA 1686

```

RESULT 2
AK082367
LOCUS
DEFINITION
Mus musculus 0 day neonate cerebellum cDNA, RIKEN full-length
(Drosophila), full insert sequence.
ACCESSION
AK082367
VERSION
AK082367.1 GI:26100612
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1
Carninci, P., and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
JOURNAL
MEDLINE
PUBMED
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL
MEDLINE
PUBMED
20499374
11042159

REFERENCE
AUTHORS
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nihei, K., Kikunishi, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
Yoneda, Y., Iehikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
RIKEN Integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE
PUBMED
120530913
11076861

REFERENCE
AUTHORS
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

REFERENCE
AUTHORS
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL
REFERENCE
AUTHORS
6 (bases 1 to 2289)
Aachi, U., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, T.,
Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
Koye, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ono, M., Ohnato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Miyamatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

COMMENT
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
location/Qualifiers

FEATURES
source
1..2289
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:C230040009"
/db_xref="taxon:10090"
/clone="C230040009"
/clone_type="cDNA library"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
152..1568
/note="naked cuticle 1 homolog (Drosophila)"
[MGD:MG12135954, GB|NM_027280, evidence: BLASTN, 98%,
match=1636]
misc_feature
match=1636

ORIGIN
Alignment Scores:
Pred. No.: 7.73e-166 Length: 2289

Score: 2179.50 Matches: 410
Percent Similarity: 90.87% Conservative: 18
Best Local Similarity: 87.05% Mismatches: 41
Query Match: 87.21% Indels: 2
DB: 3 Gaps: 1
US-09-993-966-7 (1-470) x AK082367 (1-2289)

QY 1 MetGlyLeuLeuHisSerIleProAlaAlaValCysIleArgArgIleSerProGluGly 20
Db 152 ATGGGGAACTTCACTCGAAGCCGGCCGCTGTCAGAGCGCAGAGGAGACCCGGAAAGT 211

QY 21 ArgSerPheAlaValSerAlaAlaIleTPAlaArgIleGlyIleGluGluTppIleGlyArg 40
Db 212 GACAGCTTGGCTTAAGCGCTGCTGGCAAGAAAGGCAATTCAGAGGTGATCGGAGG 271

QY 41 GlArgCysProGlyGlyValIleSerGlyProArgGlyLeuArgPheLeuAlaGlyTrrIleGly 60
Db 272 CAGCGCTGTCAGAGCAGCGCTCAGAGACCCGTCAGCTGATGGCAGGCACTTGGT 331

QY ArgSerThrArgGlyLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80
Db 332 CGAGGCACCTCGGAACTCGTGGGTGACACTTACAGAGCGCTCGGTGAGAGAGAG 391

QY 81 AspAspPheArgLeuGluValAlaLeuProProGluLeuYerThrAspGlyLeuGlySerGly 100
Db 392 GACGACTTCCCTTGAAGAGTGGCCCTGCGCCGCAAGAAATTACAGCTTAAAGTGA 451

QY 101 AspGlyLeuYerMetGluArgValIleSerGlyProCysProGlySerIleYerGlyLeuLeu 120
Db 452 GATGAGAAAGAAATGAGAGACATGAGCAAGCACTCGGCAAGGCTTCAGAAAGCACTAAG 511

QY 121 PheGluGluLeuGluCysAspValIleSerMetGluGluAspSerArg-GluGluTrrPhe 140
Db 512 TTGAAGAGCTGAGTGTATGCTCTGTGAGAGAGAGACCGGCAAGTGAAGTGA 571

QY 140 eThrIleYrAspPheAspAsnAsnGlyValIleThrArgGluAspIleThrSerLeuLeu 160
Db 572 CACTCATATGACTTGCACAAATGCAAGAGAGCCCGAGAGACCTTACCAAGCTTGC 631

QY 160 uHlStrIleTyrGlyValValAspSerSerValAsnHisSerProThrSerSerIleYer 180
Db 632 GCATACCATTTAAGAGTGTATCTCTGTGAAACATTTCCCCCAATCAAGCAAGAC 691

QY 180 cLeuArgValIleLeuThrValAlaProAspGlySerGluSerIleYerValIleVal 200
Db 692 ACTGGGGTGAAGCTCAGCGTGGCTCTGACGGAGCAGAGTGAAGAGAGCGTCTTT 751

QY 200 laengAlaAspLeuGluSerAlaArgProAlaGluThrIleYerProThrGluAspLe 220
Db 752 CAACCATTCGATCTGACAGCAAGAGCCCGCCAGAGACCAAAACCCCTAGAGAGCT 811

QY 220 uArgSerTrrGlyLeuLeuGluArgAlaProLeuArgPheGluGlyAspSerArgLeuGlu 240
Db 812 GCGTGGCTGGAGAAAGAGAGAGAGACCACTCAGGTTCCAGGGTACAGCACTGGA 871

QY 240 uGlnSerGlyCysTrrHisIleCysValAspGluAsnIleGluArgArgAsnIleTrrIle 260
Db 872 GCAGCCAGACTGCTTACCACTGCGGTGATGAGAAATTGAGAGGAAACCACTACCT 931

QY 260 uAspLeuAlaGlyTrrIleGluAsnTrrThrSerGluPheGlyProGlySerProSerValAl 280
Db 932 AGACCTGGCGGGATGAGAAACACAGCTCTGATTGACCGGAGATCCCTTCGGTGGC 991

QY 280 aglnIleSerGluLeuProProArgThrSerAspProThrArgSerArgSerIleGluPr 300
Db 992 CCAAGATCAGAGCTGCCCTCGAATCTCAACCCCACTCGCTCTGCCCAAGAGCC 1051

QY 300 ogIuAlaIleHisIleProHisArgIleYerProGluGlyValAspProAlaSerPheHisP 320
Db 1052 AGAAGCTGCCCAATCCACACCGAGAGCCCAAGGTGTGAGACCGAGGCTCTTCCACCT 1111

QY 320 eLeuAspThrProIleAlaIleValSerGluLeuGluGluArgIleYerGluAs 340

```

Db      1112 CTTGACACCCCACTTTCACAGGATCAGAGCTCCACAGCGCTCCGGGGCACTCAGGA 1171
Qy      340  pGlySerLyshIshPheValArgSerProLySAlaGlyLySersValGlyValGlyH 360
Db      1172 TGGGAGCAGACACTTGTGTGAGGTCTCCCAAGGCCCAAGGAGAACTGGGTATGGGCCA 1231
Qy      360  sValAlaArgGlyValArgSnhLySProPoleGlyProAlaIleProAlaValAspPr 380
Db      1232 CGGGGGCAGAGGTCAAGAACAGCTCCACTGTATCCACCACTACTGTCTCC 1291
Qy      380  oSerAlaHISleuAlaIleAserProAlaLeuLeuProSerLeuAlaProLeuGlyHISly 400
Db      1292 CTCTGCCCACTCTGGCCACAGCCAGCCCTTCTCCCACTCTGCACTCCGGGGCA 1351
Qy      400  sLySHISLySHISArgAlaLySArgISerGInGlyCyArGrgLyLeuGInAlaPole 420
Db      1352 GAAACACACAGCATCGAGCCAGAGGAGCCAGCGAGCTGCCGGGGCTGCAAGGCCCT 1411
Qy      420  uAlaSerGlyGly---ProValLeuGlyArgGlyHISleuArgGlyLeuProAlaLeuVa 439
Db      1412 GGCTGCAGAGAGCTCCACCGTCATGGGGCGGAGCAGGTAGGGAGCTGCTGCTGTGT 1471
Qy      439  lValTyGlySerGInAlaGlyGInProValGInArgHISGlyHISHISHISG 459
Db      1472 GGTGTACAGAGCCAGGCTGGGCAAGCCGTCACAGAGCACGAAACACATCACCA 1531
Qy      459  uHISHISHISHTyTHISHSIshPetyrGIn 469
Db      1532 ACATCACACCATTCACCACTTCTATCAG 1562

RESULT 3
AK076164
LOCUS      2471 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus 14, 17 days embryo head cDNA, RIKEN full-length
              enriched library, clone:330401D10 product:naked cuticle 1 homolog
              (Drosophila), full insert sequence.
ACCESSION  AK076164
VERSION     AK076164.1 GI:26096648
KEYWORDS   HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE   1
  AUTHORS   Carninci, P. and Hayashizaki, Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
  PUBMED    10349636

REFERENCE   2
  AUTHORS   Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
            Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
  PUBMED    11042159

REFERENCE   3
  AUTHORS   Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
            Kono, H., Akiyama, D., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
            Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
            Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
            Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Obara, E., Matsumi, M.,
            Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
            Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multipillarary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
  PUBMED    11076861

REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the

```

```

TITLE      FANTOM Consortium.
JOURNAL     Functional annotation of a full-length mouse cDNA collection
REFERENCE   Nature 409, 685-690 (2001)
AUTHORS     5
            The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
            Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
            Nature 420, 563-573 (2002)
            6 (bases 1 to 2471)
            Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
            Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
            Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
            Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Konda, M.,
            Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
            Nakamura, M., Nishi, K., Nomura, K., Numata, R., Ono, M., Ohnato, N.,
            Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
            Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
            Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
            Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
            Muramatsu, M., and Hayashizaki, Y.

TITLE      Direct Submission
JOURNAL     Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
            URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
            Fax: 81-45-503-9216)

COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL: http://genome.gsc.riken.jp/
            URL: http://fantom.gsc.riken.jp/.
            Location/Qualifiers
            source
            1..2471
               /organism="Mus musculus"
               /mol_type="mRNA"
               /db_xref="GenBank"
               /strain="C57BL/6J"
               /db_xref="FANTOM,DB:330401D10"
               /db_xref="taxon:10090"
               /clone="3230401D10"
               /tissue_type="head"
               /clone_lib="RIKEN full-length enriched mouse cDNA library"
               /dev_stage="14, 17 days embryo"
               155..1571
               /note="naked cuticle 1 homolog (Drosophila)
               (MGI:2135954, GB|NM_027280, evidence: BLASTN, 98%,
               match=1658)
               putative"

FEATURES             misc_feature
ORIGIN
Alignment Scores:
Pred. NO.:          6.75e-159
Score:              2094.50
Percent Similarity: 89.41%
Best Local Similarity: 84.75%
Query Match:        83.81%
DB:                  3
US-09-993-966-7 (1-470) x AK076164 (1-2471)
Qy      1  MecGlyLySleuHISserLySProAlaAlaValCySlySArgArgGlySerProGlyGly 20
Db      155 ATGGGGAACCTTCACTCGAAGCCGGCCGCTGTGCAAGCGAGGAGGAGCCCGGAAGGT 214
Qy      21  ApsSerPheAlaValSerAlaAlaTTPAlaArgLySgLyTleGlyGlyTTPAlaGlyArg 40
Db      215 GACAGCTTGTCTGTAAAGCGCTGTGGCAAGGAGCAATGACGAGGTGATCGGAGG 274

```

QY 41 GlnArgCysProGlyValSerGlyProArgGlnLeuAlaGlyThrIleGly 60
 Db 275 CAGGCTGTCCAGGACCGATATCAGACCCCGTCAGCTGATGGCAGGCGCTGTGT 334
 QY 61 ArgSerThrArgGlnLeuValGlyAspValLeuArgSerThrLeuSerGlnGluGlu 80
 Db 335 CGAGGCACTGGGAACTCTGGGTGACACTTCTACAGAGCTCTCGGTGAGGAGGAG 394
 QY 81 AspArgPheArgLeuGlnValAlaLeuProGlnGlyThrAspGlyLeuGlySerGly 100
 Db 395 GAGCACTTCCCTCAACAAGTGGCCCTGCGCTGAAAGATTGACAGCTAGATGATGGA 454
 QY 101 AspGlnGlySerMetGlnArgValSerGluProCysProGlySerGlyGlnLeuGly 120
 Db 455 GATGACAGAGAAATGAGAGAGAGAGTGCCTGCGCCAGCGCTCCAGAAAGCACTCAG 514
 QY 121 PheGlnGlnLeuGlnCysAspValSerMetGlnGlnAspSerArgGlnGlnThrPhe 140
 Db 515 TTGGAACAGCTGCACTGTGATGTCTGTGAGAGAGAGAGAGCGGCAAGAGTGAAGCTTTC 574
 QY 141 ThrLeuThrAspPheAspAsnAsnGlyValValThrArgGlnAspIleThrSerLeu 160
 Db 575 ACTCTATATGACTTCGACAAACATGGCAAGTGAACCGTGAAGACATTACAGCTGCTG 634
 QY 161 HisThrIleTyrGlnValAlaAspSerSerValAsnHisSerProThrSerSerLeu 180
 Db 635 CATACCTATATGAAAGTGTGATCTCTGTGAAACATTCCTCCCAATCAGCAAGACA 694
 QY 181 LeuArgValLys-LeuThrValAlaProAspGlySerGlnSerTyrAspSerValLeu 200
 Db 695 CTGGGGGTGAAGCTCAACCGGTCTCTGAGGAGGACACGTAAGAGAGAGCGTCTTT 754
 QY 200 LaeGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAsp 220
 Db 755 CAACCATACCGATTCGACAGACACAGAGCCCGCGGACGA-ACCAACCCCGTGAAGAGCT 813
 QY 220 ValArgSerTyrGlnLysGlnAlaGlnAlaProLeuArgPheGlnGlyAspSerArgLeu 240
 Db 814 GCCTGGGTGGAGAAAGAGCAGCGCCCACTCAGGTTTCAGGGTACACCCCACTGGA 873
 QY 240 uGlnSerGlyCysTyrHisHisCysValAspGlnAsnIleGlnLysArgAsnHisTyr 260
 Db 874 GCGAGCCAGACTGTACACCATTCGTGATGAGAAACATGAGAGGAAACCATCACTCT 933
 QY 260 uAspLeuAlaGlyIleGlnLysTyrThrSerGlnPheGlyProGlySerProSerValAla 280
 Db 934 AGACCTGGCGGGATAGAAACTACACTCTCACTTTGGACGGGATCCCTTCGTGGC 993
 QY 280 ArgLysSerGlnLeuProProArgThrSerAsnProThrArgSerArgSerHisGlu 300
 Db 994 CCAAGATCAGAGGTGCCCTCGAATCTCCACCCCACTGCTCTGCTCCCAAGACC 1053
 QY 300 OGlnAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHis 320
 Db 1054 AGAAGCTGCCCACTCCACACCGAGGCCCCAAGAGGTGAGCCAGAGCTCTCCACT 1113
 QY 320 eLeuAspThrProIleAlaLysValSerGlnLeuGlnGlnArgLeuArgGlyThrGln 340
 Db 1114 CTTTGACACCCCTTTTCCCAAGCATCAGACTCCAGCAAGCGCTCCGGGCACTCAGGA 1173
 QY 340 pGlySerLeuHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValGly 360
 Db 1174 TGGGAGCAACCACTTTGTGAGGTCCCAAGGCCCCAAGGCGAAGAACTGGGTATGGCCA 1233
 QY 360 eValAlaArgGlyValArgAsnLysProProLeuGlyProAlaIleProAlaValSer 380
 Db 1234 CGGGGCGAGAGGTGCAAGCAAGCAAGCTCCACTGTATCCACCACTATCTGCTCC 1293
 QY 380 oSerAlaHisLeuAlaLysSerProAlaLeuLeuProSerLeuAlaProLeuGlyHis 400
 Db 1294 CTCTGCCCATCTGGCCACACGACCCCTTCTCCCACTGGCAACCCCTGGGGGACAA 1353

QY 400 eLysHisLeuArgAlaLysGlnSerGlnGlnGlyCysArgGlyLeuGlnAlaProLe 420
 Db 1354 GAACACAGACATCAGACCAAGAGAGCCAGGCGAGCTGCCGGGCTGGCAGGGCCCT 1413
 QY 420 uAlaSerGlyGly---ProValLeuGlnArgGlnHisLeuArgGlnLeu-ProAlaLeu 439
 Db 1414 GGCTGCAGAGAGGCTCCACCGTCACTGGGCGGAGAGAGTGAAGGAGCTGGCGTGG 1473
 QY 439 aValTyrGlnSerGlnAlaGlyGlnProValGlnArgHisGlnHisHisHisG 459
 Db 1474 TGGTGTACGAGAGGACAGCTGGGCGAGCGCTCCAGAGACAGAACCATCACCACG 1533
 QY 459 IuhHisHisHisTyrHisHisPheTyrIle 469
 Db 1534 AACATCACACCATTCACACCACTTCTATCAG 1565
 RESULT 4
 AK013241
 LOCUS
 DEFINITION
 Mus musculus 10, 11 days embryo whole body cDNA, RIKEN full-length
 (Drosophila), full insert sequence.
 ACCESSION
 AK013241 GI:12850482
 VERSION
 AK013241.1
 KEYWORDS
 HTC; CAP trapper.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE
 1 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishii, K., Kiteunai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishure, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishwagi, K.,
 Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Onara, E., Watanaka, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 4 The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5 The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 1466)
 Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,
 Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
 Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hizumoto, K.,
 Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
 Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,

GGCAGCAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC library."

ORIGIN

Alignment Scores:

Pred. No.:	3,26e-108	Length:	931
Score:	1462.00	Matches:	270
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	58.50%	Indels:	0
DB:	5	Gaps:	0

US-09-993-966-7 (1-470) x BQ653673 (1-931)

```

Qy 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeu 220
    |||||
Db 3 AATCAGGCTACTGCGAGAGCGAGGCCCGGAGCAGAGACCAAGCCCACTGAGGACCTG 62
    |||||
Qy 221 ArgSerTrpGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
    |||||
Db 63 CGGAGCTGGGAGAAAGAGCAGCGAGCCCGCTCAGGTTCCAGGGGTGACAGCCGCTGAG 122
    |||||
Qy 241 GlnSerGlyCysTrpTrpHisCysValAspGluAsnIleGluArgGlnHisTrpLeu 260
    |||||
Db 123 CAGCTGGCTGCTACCACTTGCGTATGATGAGAACATCAAGAGGAAACCACTACTCTTA 182
    |||||
Qy 261 AspLeuAlaGlyIleGluAsnTrpThrSerGlnPheGlyProGlySerProSerValAla 280
    |||||
Db 183 GATCTGCCCGGATGAAAATCAACGCTCCCAATTTGGGCTGCTCCCTCCGTGCGC 242
    |||||
Qy 281 GlnLysSerGluLeuProProArgTrpThrSerAsnProThrArgSerArgSerHisGluPro 300
    |||||
Db 243 CAGAGTCAAGAACTGCCCCCGGACCTCCCACTCCCATCTCGATCTCGCTCCCATGAGCCG 302
    |||||
Qy 301 GlnAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPhe 320
    |||||
Db 303 GAAGCCATTCACATCCACACCGAAAGCCCAAGCGGTGAGCCGCGCTCCCTCCACTTC 362
    |||||
Qy 321 LeuAspThrProIleAlaLysValSerGluLeuGlnGlnArgLeuArgGlyThrGlnAsp 340
    |||||
Db 363 CTGAGCAACCCCAATCGCCCAAGGTCTCAGAGCTCCAGAGCGGCTCCGGGGACCCAGAGAC 422
    |||||
Qy 341 GlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValGlyHis 360
    |||||
Db 423 GGGAGCAAGCACTTGTGAGGTCCCCCAAGGCCCAAGGCAAGAGTGTGGGTGGGCAC 482
    |||||
Qy 361 ValAlaArgGlyValaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
    |||||
Db 483 GTGGCCAGAGGGGAGAAACMAAGCCCTCTGGGACCCGCACTCCGCTGCTCCGCC 542
    |||||
Qy 381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
    |||||
Db 543 TCGGCCAATCGGTGTGCGAGCGCCGCTCTCTCCCTCTAGGCCCTCCCTCGGGACAG 602
    |||||
Qy 401 LysHisLysHisArgAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
    |||||
Db 603 AAGCACAAGCACCAGGCAAGAGAGCCAGAGGGGTGCGGGGGCTGAGAGGACCACTG 662
    |||||
Qy 421 AlaSerGlyLysProValLeuGlyArgGluHisIleLeuArgGlyLeuProAlaLeuValVal 440
    |||||
Db 663 GCTCAGAGTGGCCCTGTCTGGGGCGGAGACCTTGCGGAGCTGCGCGCTTGTGTGTG 722
    |||||
Qy 441 TyrGluSerGlnAlaGlyGlnProValGlnArgHisGlnHisHisHisHisHisHis 460
    |||||
Db 723 TATGAGGCCAGGCGCGGAGCCGCTCCAGAGATGAGACCAACCAACCAATGAACAT 782
    |||||
Qy 461 HisHisHisTrpHisHisPheTrpGlnThr 470
    |||||
Db 783 CACACCATTAACCACTTCTACAGACA 812
    |||||

```

RESULT 7

BQ645656

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

TITLES

AUTHORS

JOURNAL

COMMENT

BQ645656 921 bp mRNA linear EST 15-JUN-2002
 AGENCOURT_8355700 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6285268
 5', mRNA sequence.
 BQ645656.1 GI:21769828
 EST.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 921)
 NIH-MGC http://mgs.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-x@mail.nih.gov
 Tissue Procurement: CGAP (Stanford)
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 plate: LNCM2484 row: d column: 05
 High quality sequence start: 16
 High quality sequence stop: 685.
 Location/Qualifiers
 1. 921
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6285268"
 /tissue_type="hepatocellular carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_1ib="NIH MGC 100"
 /note="Organ: liver; Vector: pORF7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dr priming. Directionally cloned
 into BcoRI/XhoI sites using the following 5' adaptor:
 GGCACAG(G). Size-selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the Laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH-MGC
 Library."

FEATURES

source

ORIGIN

Alignment Scores:

Pred. No.:	4.13e-106	Length:	921
Score:	1436.00 <td>Matches:</td> <td>266 </td>	Matches:	266
Percent Similarity:	98.89% <td>Conservative:</td> <td>1 </td>	Conservative:	1
Best Local Similarity:	98.52% <td>Mismatches:</td> <td>3 </td>	Mismatches:	3
Query Match:	57.46% <td>Indels:</td> <td>0 </td>	Indels:	0
DB:	5	Gaps:	0

US-09-993-966-7 (1-470) x BQ645656 (1-921)

```

Qy 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeu 220
    |||||
Db 22 CATCAGGCTACTGCGAGAGCGAGGCCCGGAGCAGAGACCAAGCCCACTGAGGACCTG 81
    |||||
Qy 221 ArgSerTrpGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
    |||||
Db 82 CGGAGCTGGGAGAAAGAGCAGCGAGCCCGCTCAGGTTCCAGGGGTGACAGCCGCTGAG 141
    |||||
Qy 241 GlnSerGlyCysTrpTrpHisCysValAspGluAsnIleGluArgGlnHisTrpLeu 260
    |||||
Db 142 CAGCTGGCTGCTACCACTTGCGTATGATGAGAACATCAAGAGGAAACCACTACTCTTA 201
    |||||
Qy 261 AspLeuAlaGlyIleGluAsnTrpThrSerGlnPheGlyProGlySerProSerValAla 280
    |||||

```


Db 202 GATCTGCGCGGAGTAAAGAACTACAGTCCCAATTTGGGCTGGCTCCCTTCGTCGCC 261
Qy 281 GlnlybSerGluLeuProProAlaGThrSerAsnProThraArgSerArgSerHisGluPro 300
Db 262 CAGAAAGTCAGAACTGCCCCCGGACCTCCCAATCCCATGATCTCGCTCCCAAGAACCG 321
Qy 301 GlnAlaIleHisIleProHisArgGlybProGlnGlyAlaAspProAlaSerPheHisPhe 320
Db 322 GAAAGCCATCCACATCCCAACCCGAAAGCCCAAGGCGGTGAACCGGCTTCCTCCACTTC 381
Qy 321 LeuAspThrProIleAlaValSerGluLeuGlnGlnIleArgLeuArgGlyThrGlnAsp 340
Db 382 CTGGACACCCCATTCGGCAAGTCTCAGAGCTCCAGAGCGGCTCCGGGGCAACCCAGAC 441
Qy 341 GlySerIleHisPheValArgSerProGlyAlaGlnGlybSerValGlyValGlyHis 360
Db 442 GGGAGCAAGCACTTTGTGAGGTCCCCCAAGGCCCAAGGCAAGAGTGTGGGTGGCCAC 501
Qy 361 ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
Db 502 GTGGCCAGAGGGGCAAGAAACAAGCCCTCTGGAGCCCGCATCCCTGGCGGTCTCCCC 561
Qy 381 SerAlaHisLeuAlaAlaSerProAlaSerLeuLeuProSerLeuAlaProLeuGlyHis 400
Db 562 TCGGCCACCTGGCTGGCAAGCCCGCCCTCTCCCTCCCTAGCCCCCTCGGGCAAG 621
Qy 401 LysHisIleHisArgAlaIleGlybSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
Db 622 AACCAACAAGCAGCAGGCAAGAGAGCACAAGGCTGCGGGGCTGACAGCACCAC 681
Qy 421 AlaSerGlyGlyProValLeuGlyArgGluHisIleuArgGlyLeuProAlaLeuVal 440
Db 682 GCTCTCAGTGGCTGCTGCTGGGGGGGAGCACTGGGAGCGCCCTGTGGTG 741
Qy 441 TyrGluSerGlnAlaGlyGlnProValGlnArgHisGluHisIleHisIleHisGluHis 460
Db 742 TATGAGAGCCAGGCGCGGAGCGGCTCCACAGACATAGACACCAACCATGAACAT 801
Qy 461 HisHisIleTyrHisIleHisPheTyrGlnThr 470
Db 802 CACCACCATTAACCACTCATCAAGACA 831
RESULT 8
LOCUS BQ064678 990 bp mRNA linear EST 02-Apr-2002
DEFINITION AGENCOURT 6853565 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5926427
5', mRNA sequence.
ACCESSION BQ064678
VERSION BQ064678.1 GI:19893537
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 990)
NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Lou Staudt
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LINC2099 row: 1 column: 12
High quality sequence stop: 670.
Location/Qualifiers
1. .990
/organism="Homo sapiens"
/mol_type="mRNA"

/db xref="taxon:9606"
/clone="IMAGE:5926427"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_99"
/note="Organ: lymph. Vector: pOT7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
library."

ORIGIN
Alignment Scores:
Pred. No.: 9,08e-105 Length: 990
Score: 1420.00 Matches: 268
Percent Similarity: 98.53% Conservative: 0
Best Local Similarity: 98.53% Mismatches: 2
Query Match: 56.82% Indels: 2
DB: Gaps: 0
US-09-993-966-7 (1-470) x BQ064678 (1-990)
Qy 201 AsnGlnAlaAspLeuGlnSerAlaArgProAlaGlnThrLysProThrGluAspLeu 220
Db 3 AATCAGGCTGACCTGCAGAGCGCAAGGCCCGGAGAGACCAAGCCCATGAGACCTG 62
Qy 221 ArgSerTrpGlnLysbSerGlnArgAlaProLeuArgPheGlnIleLysAspSerArgLeuGlu 240
Db 63 CGAGGCTGGAGAGAAAGCAGCAGCGCCGCTCAGGTTCCAGGGTGAACCCGCTGGAG 122
Qy 241 GlnSerGlyCysTyrHisIleCysValAspGluAsnIleGluArgArgAsnIleSerLeu 260
Db 123 CAGTGTGGCTGCTACCACTTCGCTGATGAAACATCGAGAGAGAAACCACTACTTA 182
Qy 261 AspLeuAlaGlyIleGluAsnTyrTrpSerGlnPheGlyProGlySerProSerValAla 280
Db 183 GATCTGCGCGGAGTAAAGAACTACAGTCCCAATTTGGGCTGGCTCCCTTCGTCGCC 242
Qy 281 GlnlybSerGluLeuProProAlaGThrSerAsnProThraArgSerArgSerHisGluPro 300
Db 243 CAGAAAGTCAGAACTGCCCCCGGACCTCCCAATCCCATGATCTCGCTCCCAAGAACCG 302
Qy 301 GlnAlaIleHisIleProHisArgGlybProGlnGlyAlaAspProAlaSerPheHisPhe 320
Db 303 GAAAGCCATCCACATCCCAACCCGAAAGCCCAAGGCGGTGAACCGGCTTCCTCCACTTC 362
Qy 321 LeuAspThrProIleAlaValSerGluLeuGlnGlnIleArgLeuArgGlyThrGlnAsp 340
Db 363 CTGGACACCCCATTCGGCAAGTCTCAGAGCTCCAGACACGCTCGGGGCAACCCAGAC 422
Qy 341 GlySerIleHisPheValArgSerProGlyAlaGlnGlybSerValGlyValGlyHis 360
Db 423 GGGAGCAAGCACTTTGTGAGGTCCCCCAAGGCCCAAGGCTGGGTGTGGGCGCAC 482
Qy 361 ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
Db 483 GTGGCCAGAGGGGCAAGAAACAAGCCCTCTGGAGCCCGCATTCCTCGGTGTCCCC 542
Qy 381 SerAlaHisLeuAlaAlaSerProAlaSerLeuLeuProSerLeuAlaProLeuGlyHis 400
Db 543 TCGGCCACCTGGCTGGCAAGCCCGCCCTCTCCCTCCCTAGCCCCCTCGGGCAAG 602
Qy 401 LysHisIleHisArgAlaIleGlybSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
Db 603 AACCAACAAGCAGCAGGCAAGAGAGCACAAGGCTGCGGGGCTTGAAGGACCACTG 662
Qy 421 AlaSerGlyGlyProValLeuGlyArgGluHisIleuArgGlyLeuProAlaLeuVal 440
Db 663 GCTCTCAGTGGCTGCTGCTGGGGGGGAGCACTCGGGAGGCTGCCCGCTTGTGGTG 722

Oy		441	TtYgUserGlnAa-GlyglInProValGlNArghIselUnIshIShISHIGHLNT 460
Dd		723	TATAGAGACCACGGCCGGGCAGCCCGGTCTCAAGAAGATTGAAGCACCAACCATGAAACA 782
Oy		460	sHIShISHISTyrHis-HisPhETyrGIIntur 470
Dd		783	TCACCACATTACCAACCACTTCTCAGAGACA 814
RESULT 9			
LOCUS	B0644360		
DEFINITION	AGENSCOURT 8351368 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:628691.7		
ACCESSION	B0644360		
VERSION	B0644360.1	GI:21768532	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. NIH-MGC http://mgc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strusberg, Ph.D. Email: cgabds-remail.nih.gov Tissue Procurement: CGAP (Stanford) cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L1CM2488 row: h column: 22 High quality sequence stop: 622. Location/Qualifiers 1..895 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:628691.7" /tissue.type="hepatocellular carcinoma, cell line" /lab.host="DH10B (phage-resistant)" /clone_id="NIH_MGC_100" /note="Organ: liver; Vector: pOTB7, Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directional cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."		
ORIGIN			
Alignment Scores:			
Pred. No.:	5,76e-103	Length:	895
Score:	1397.00	Matches:	261
Percent Similarity:	98.13%	Conservative:	1
Best Local Similarity:	97.75%	Mismatches:	5
Query Match:	55.90%	Indels:	0
DB:	5	Gaps:	0
US-09-993-966-7 (1-470) x B0644360 (1-895)			
Oy	201	AenGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIlyProThrcyluAspleu 220 ::	
Dd	4	CATCAGGCTGACTCTCAAGAGCCCAAGCCCCGAGACAGAACCAACCCACTGAGGAACTTG 63 	
Oy	221	ArgSetrPGJuLyalyeGINarGLaProlouargpheginglYaSPSeTargleuGu 240 ::	
Dd	64	CGAGCTCGGAGAAAGAGGAGCCCCTCAGATTCCAGGGTGA CAGCCGCTTGAG 123 ::	

QY	241	GlusbrgTgCyetyrThshisCyevaValAspGlnasnl1eglunrgrgAspnh1stYrLieu	260
Db	124	CAGTCTGGCTGTCAACCACTTGCGTGAATGAGAACATCAAGAGAAACCACTACTTA	183
QY	261	AspLeuAagly11egluAenTYrThSerGlnPheGlyProGlySerProSerValAla	280
Db	184	GATCTCGCGGGGATAGAAAACCTACAGCTCCCAATTGGGCTGGCTCCCTCGGTGGCC	243
QY	281	GlnYsSerGluLeuProProAlaGThSerSerAspProThrArgSerAArgSerH1eglunPro	300
Db	244	CAGAAgTCAGAACTGCCCCCCCCCGCACTCCCAATCCCACTCGATCTCCCTCCCATGAGCCG	303
QY	301	GluAla11eh1st1leProH1sarGlyeProGlnGlyValAspProAlaSerPheH1sPhe	320
Db	304	GAAgCATTCACATCTCCACACCGAAAGCCCCCAAGGCGTGACCCCGCTCTTCCACTTC	363
QY	321	LeuAspThrPro11eal1alyeValSerGluLeuGlnGlnArgLeuAArgGlyThrGlnAsp	340
Db	364	CTTGACACCCCAATCGCCCAAGGTCTCAAGGCTCAGACACGGCTCGGGGCACCCAAGAC	423
QY	341	GlySerTlysh1sPheValArgSerProTlyeAl1aglnGlyYsSerVal1GlyValGlyH1s	360
Db	424	GGGAGCAAGCACTTGGAGGTCCCCCAAGGCCCAAGGCAAGATGTGGGTGGGGCAC	483
QY	361	Val11ArgrGlyAlaArgAsnYsProProLeuGlyProAla11eProAlaValAserPro	380
Db	484	GTGGCCAGAGGGGCAAGAAACAAAGCCCCCTCTGGGAACCCGCATCCTTGCGGTGTCCCCC	543
QY	381	Ser11Ah1sLeuAla1AserProAlaLeuLeuProSerLeuAlaProLeuGlyH1sYs	400
Db	544	TCCGGCCACCTGGGTGCCAGGCCCGGCGCTCTCCCTCCCTAGGCCCTCGGGCAAG	603
QY	401	LysH1slysh1sArGAl1alyeGluSerGlnGlnGlyCySarGlyLeuGlnAlaProLeu	420
Db	604	AAGCAACAAGCAAGCCAGCCAAAGAGCAGAGGGGCTGGCCGGGCTGTCAAGCAACCACTG	663
QY	421	AlaSerGlyGlyProVal1LeuG1AArgGlyH1sPheArgGluLeuProAlaLeuVal1	440
Db	664	GCTTCAGGTGGCCCTGTCTCTGGGGCGGAGCACTCGCGGAGCTGCCCCCTTGGTGTGTG	723
QY	441	TyrGluSerGlnAlaGlyGlnProValGlnArGh1sGlnH1sSh1sH1sH1sH1sH1s	460
Db	724	TATGAGAGCCAGGCGGGGACAGCGGTCAAGAGCATGAGCAACCCCAACATGAGCAT	783
QY	461	HisH1sH1sTlyrH1sH1sPhe	467
Db	784	CACCACTTTACACACACTTT	804
RESUR.T 10			
B0644956			
LOCUS			
DEFINITION	B0644956	928 bp	mRNA
ACCESSION	AGNCOURT_8511562	NIH_MGC_100	Homo sapiens cDNA clone IMAGE:6297032
VERSION	B0644956		5', mRNA sequence.
KEYWORDS	B0644956.1	GI:21769128	EST.
SOURCE			
ORGANISM			
	Homo sapiens		(human)
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE			
AUTHORS	1 (bases 1 to 928)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgsabds-remail.nih.gov		
	Tissue Procurement: CGAP (Stanford)		
	cDNA Library Preparation: Rubin Laboratory		
	DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LNLB)		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNLB at:		

http://image.jnl.gov
 Plate: LICM2505 row: n column: 09
 High quality sequence start: 17
 High quality sequence stop: 551.

FEATURES

source

1. 928
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6297032"
 /issue_type="hepatocellular carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC_100"
 /note="Organ: liver; Vector: pOTB7, Site 1: XhoI, Site 2: EcoRI, cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."

ORIGIN

Alignment Scores:

Pred. No.: 9,99e-102 Length: 928
 Score: 1382.00 Matches: 261
 Percent Similarity: 96.68% Conservative: 1
 Best Local Similarity: 96.31% Mismatches: 7
 Query Match: 55.30% Indels: 2
 DB: 5 Gaps: 0

US-09-993-966-7 (1-470) x BQ644956 (1-928)

QY 201 AaGlnAlaApLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeu 220
 Db 18 CATCAGGCTGACTGCGAGAGCGCAGCGCCCGAGCAGAGCCAGCCACTGAGGACTTG 77
 QY 221 ArgSerTrpGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
 Db 78 CGAGAGCTGGAGAGAGAGAGAGAGCGCCGCTGAGGTTCAGAGGTGACACCGCTGGAG 137
 QY 241 GlnSerGlyCysTrpTrpHisHisCysValAspGluAsnIleGluArgArgAsnTrpLeu 260
 Db 138 CACTCTGGCTGCTACCACTTCGTAGATGAGAACATCCAGAGGAGAAACCACTACTTA 197
 QY 261 AspLeuAlaGlyIleGluAsnTrpTrpSerGlnPheGlyProGlySerProSerValAla 280
 Db 198 GATCTGCGCGGATGAGAACTACACCTCCCAATTGAGCTGCTCCCTTCGTTGCC 257
 QY 281 GlnLysSerGluLeuProProArgTrpSerAsnProThrArgSerArgSerHisGluPro 300
 Db 258 CAGAAGTCAGAACTGCCCCCGGACCTCCAACTCCCATGATTCGCTCCCATGAGCG 317
 QY 301 GluAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPhe 320
 Db 318 GAAGCCATCCACATCCACACCAAGAGCCCAAGAGCGCTGAGACCGGCTCTTCACCTTC 377
 QY 321 LeuAspThrProIleAlaLysValSerGluLeuGlnArgLeuArgGlyThrGlnAsp 340
 Db 378 CTGGACACCCCAATCGCCAGAGGTCTCAGAGCTCCAGAGAGGCTCCGGGACACCCAGAC 437
 QY 341 GlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValGlnHis 360
 Db 438 GGGAGCAAGCACTTTGAGGTCTCCCAAGGCCCAAGGCAAGAGTGGGTGGGCGCAC 497
 QY 361 ValAlaArgGlyValArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
 Db 498 GTGGCCAGAGGGCCAGAAACAGAGCCCTCTGGAGCCCGCCATCCCTGGGTGTCCCC 557
 QY 381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlnHisLys 400
 Db 558 TCGGCCCACTGGCTGCGAGCCGAGNCCTCTCCCTCCAGCCCTTCGGGACACAG 617

QY 401 LysHisLysHisArgAlaLysGlnSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
 Db 618 AACACAG 677
 QY 421 AlaSerGlyGlyProValLeuArgLysArgGluHisIleLeuArgGluLeuProAlaLeu-Val 440
 Db 678 GCTTCAGGTGCGCTCTGCTGGAGCGGAGACCTTCGGAGAGCTGCGGCTGCTGGTGG 737
 QY 440 LTYGtLysSerGlnAlaGlyLysProValGlnArgHisGluHisIleHisHisHisIleGlu 460
 Db 738 GTATGAG 797
 QY 460 IshHisHisIleTrpHisHisPheTrpGln 469
 Db 798 ATCCACACCATTCACACCATTCATCCAG 826

RESULT 11

LOCUS AY412100 888 bp DNA linear GSS 16-DEC-2003
 DEFINITION Mus musculus NKD1 gene, VIRUTAL TRANSCRIPT, partial sequence,
 ACCESSION AY412100
 VERSION AY412100.1 GI:39768065
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus-
 1 (bases 1 to 888)
 Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
 Todd, M.A., Tanenbaum, D.M., Civeillo, D.R., Lu, F., Murphy, B.,
 Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sainsky, J.J.,
 Adams, M.D. and Cargill, M.
 Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 Science 302 (5652), 1960-1963 (2003)
 14671302
 2 (bases 1 to 888)
 Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
 Todd, M.A., Tanenbaum, D.M., Civeillo, D.R., Lu, F., Murphy, B.,
 Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sainsky, J.J.,
 Adams, M.D. and Cargill, M.
 Direct Submission
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 This sequence was made by sequencing genomic exons and ordering
 them based on alignment.
 COMMENT Location/Qualifiers
 FEATURES
 source
 1. 888
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 <1..>888
 /gene="NKD1"
 /locus_tag="NCM4444"

ORIGIN

gene

Alignment Scores:
 Pred. No.: 4.18e-101 Length: 888
 Score: 1374.00 Matches: 259
 Percent Similarity: 91.22% Conservative: 11
 Best Local Similarity: 87.50% Mismatches: 26
 Query Match: 54.98% Indels: 0
 DB: 9 Gaps: 0

US-09-993-966-7 (1-470) x AY412100 (1-888)

QY 65 GluLeuValGlyAspValLeuArgAspTrpLeuSerGlnGluGlnGluAspAspPheArg 84
 Db 1 GAACGTGGGTGACACTTCTAGAGAGGCTCTCGTGGAGAGAGACGAGAGCACTTCCC 60
 QY 85 LeuGluValAlaLeuProProGlyLysTrpAspGlyLeuGlnGlySerGlyAspGlyLysLys 104

```

Db      61 CTGAGATGGCCCTGCGCCCTGAGAAATGACAGCTAGGTGTGAGATGAGAAAGA 120
Qy      105 MetGluArgValSerGluProCysProGlySerIlySerGluLeuIlySerGluLeu 124
Db      121 ATGAGAGAGATGGCGGACCTGGCCAGGCTCCCAAGAGAGAGCTCAAGTTTGAAGAGCTG 180
Qy      125 GluCysAspValSerMetGluIlyAspSerArgGlnIlyThrPheThrLeuIlyAsp 144
Db      181 CAGTGTATGTCTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Qy      145 PheAspAsnAsnGlyIlyValThrArgIlyAspIleThrSerLeuLeuIlySerIleTyr 164
Db      241 TTGAGACAACATGGCAAAAGTACCCGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Qy      165 GluValValAspSerSerValAsnHisSerProThrSerSerIlyMetLeuArgValIly 184
Db      301 GAGGTGGTTGATCTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Qy      185 LeuThrValAlaProAspGlySerGlnSerIlySerValLeuValAsnGlnAlaAsp 204
Db      361 CTGACCTGGCTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Qy      205 LeuGlnSerAlaArgProArgAlaGluThrIlySerProThrGluAspLeuArgSerTyrGlu 224
Db      421 CTGAGAGAGACAAGAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Qy      225 IlyIlySerGlnArgAlaProLeuArgPheGlnIlyAspSerArgLeuGlnIlySerIlyCys 244
Db      481 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
Qy      245 TyrHisGlyCysValAspGluAsnIleGluArgArgAsnIlySerIlyLeuAspLeuAlaGly 264
Db      541 TACCAACATTTGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Qy      265 IleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValAlaGlnIlySerGlu 284
Db      601 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Qy      285 LeuProProArgThrSerAsnProThrArgSerArgSerHisGluProGlnAlaIleHis 304
Db      661 CTGCGCCCTCGAATCTCAACCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Qy      305 IleProHisArgIlySerProGlnIlyValAspProAlaSerPheHisPheLeuAspThrPro 324
Db      721 ATCCACACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Qy      325 IleAlaIlySerGluLeuGlnIlyArgLeuArgIlyThrGlnAspGlySerIlyHis 344
Db      781 TTTCGCAAGGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
Qy      345 PheValIlySerProIlyAlaGlnIlyIlySerValGlyValGlyHis 360
Db      841 TTTCGTGAGGTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 888

RESULT 12
B0646371 1008 bp mRNA linear EST 15-JUN-2002
LOCUS B0646371
DEFINITION AGNCOURT 8493065 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6300188
5', mRNA sequence.
ACCESSION B0646371
VERSION B0646371.1 GI:21770543
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1008)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaaps-remail.nih.gov

```

```

FEATURES
    source
        1..1008
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:6300188"
        /tissue_type="hepatocellular carcinoma, cell line"
        /lab_host="DH10B (phage-resistant)"
        /clone_1ib="NIH MGC_100"
        /note="Organ: liver; Vector: pORF7; Site 1: XhoI; Site 2:
        EcoRI; cDNA made by oligo-dT priming. Directionally cloned
        into EcoRI/XhoI sites using the following 5' adaptor:
        GGCACAGAG(G). Size-selected >500bp for average insert size
        1.8kb. Library constructed by Ling Hong in the Laboratory
        of Gerald M. Rubin (University of California, Berkeley)
        using ZAP-cDNA synthesis kit (Stratagene) and Superscript
        II RT (Life Technologies). Note: this is a NIH_MGC
        library."

ORIGIN
Alignment Scores:
Pred. No.: 1,09e-93 Length: 1008
Score: 1283.50 Matches: 254
Percent Similarity: 93.45% Conservative: 3
Best Local Similarity: 92.36% Mismatches: 11
Query Match: 51.36% Indels: 7
DB: Gaps: 2

US-09-993-966-7 (1-470) x B0646371 (1-1008)
Qy      201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIlySerProThrGluAspLeu 220
Db      22 CATCAGGCTGACTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 81
Qy      221 ArgSerTyrGluIlySerGlnArgAlaProLeuArgPheGlnIlyAspSerArgLeuGlu 240
Db      82 CGAGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 141
Qy      241 GlnSerGlyCysTyrHisIlySerValAspGluAsnIleGluArgArgAsnIlySerIlyLeu 260
Db      142 CAGTGTGGCTGTACCAACATTCGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 201
Qy      261 AspLeuAlaGlyIleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValAla 280
Db      202 GATCTGCCGGAGTAGAAACTCAAGTCCCAATTGGGCTGAGCTCCCTTCCTGGAGCC 261
Qy      281 GlnIlySerGluLeuProProArgThrSerAsnProThrArgSerArgSerHisGluPro 300
Db      262 CAGAGAGCAAAAGTGGCCCCCGACCTCCCAATCCCACTGATTCCTCTCCATGAGACCG 321
Qy      301 GluAlaIleHisIleProHisArgIlySerProGlnIlyValAspProAlaSerPheHisPhe 320
Db      322 GAGGCCATCCATCCACACACGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 381
Qy      321 LeuAspThrProIleAlaIlySerGluLeuGlnIlyArgLeuArgIlyThrGlnAsp 340
Db      382 CTGAGACACCCCAATCCGCAAGTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 441
Qy      341 GlySerIlyHisPheValArgSerProIlyAlaGlnIlyIlySerValGlyValGlyHis 360
Db      442 GGGAGCAAGCACTTTGTGAGAGTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 501
Qy      361 ValAlaArgGlyAlaArgAsnIlyProProLeuGlyProAlaIle-ProAlaValSerPr 380

```

Db 502 GTGGCCAGAGGGGACAGAAACAGACCCCTCTGGAGACCCGACATCCCTCGGTGTCTCC 561

Qy 380 OSerAlaHleuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHly 400

Db 562 CTCGGCCACCTGGCTGCGAGCCGCGGCTCTCCCTCCCTAGCCCCCTCGGACAA 621

Qy 400 GlyHleuHlyHleuAlaAlaSerGlnGlnGlyCyAArgGlyLeuGlnAlaProLe 420

Db 622 GAAGCAACAGACGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 681

Qy 420 uAlaSerGlyGlyProVal-LeuGlyArgGlnHlyHleuArgGlnLeuProAlaLeu--- 438

Db 682 GGCCTCAGGTGGCTCTGTTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 741

Qy 439 ---ValValTyrgLuserGlnAlaGlyGlnPro---ValGlnArgHlyGlnHlyHleu 456

Db 742 GGGGGGGTATGAAAGCCCGAGCCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 801

Qy 457 -HisHisGlnHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 468

Db 802 CCACCATGAACATCACCCCATTTACACCACTTCTAC 838

RESULT 13

LOCUS B0652087 996 bp mRNA linear EST 15-JUL-2002

DEFINITION AGENCOURT_8207495 NIH_MGC_100 Homo sapiens CDNA clone IMAGE:6283110

ACCESSION B0652087

VERSION B0652087.1 GI:21776259

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1. (bases 1 to 996)

AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Euteheria; Primates; Catarrhini; Homnidae; Homo.

TITLE NIH-MGC http://mgc.ncl.nih.gov/

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cga@ds-remail.nih.gov

Tissue Procurement: CGAP (Stanford)

CDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov

plate: LICM2478 row: j column: 07

High quality sequence stop: 604.

FEATURES

source

1. 996

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6283110"

/tissue_type="hepatocellular carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_idb="NIH MGC 100"

/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 3.85e-91 Length: 996

Score: 1252.00 Matches: 243

Percent Similarity: 94.27% Conservative: 4

Best Local Similarity: 92.75% Mismatches: 12

Query Match: 50.10% Indels: 3

DB: 5 Gaps: 2

US-09-993-966-7 (1-470) x B0652087 (1-996)

Qy 201 AsnGlnAlaAspLeuGlnSerAlaArgProAlaGlyHlyHleuProThrGlnAspLeu 220

Db 3 AATCCGGCTTACCTGACAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 62

Qy 221 ArgSerTrpGlnLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240

Db 63 CGAGTCTGGAG 122

Qy 241 GlnSerGlyCyTrpHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 260

Db 123 CAGTCTGGCTGCTACCACTTGGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 182

Qy 261 AspLeuAlaGlyTleGlnAsnTyrrHisSerGlnPheGlyProGlySerProSerValAla 280

Db 183 GATCTCGCCGGAGTAG 242

Qy 281 GlnLysSerGlnLysProProAlaArgThrSerAsnProThrArgSerArgSerHisGlnPro 300

Db 243 CAGAGTCAAGACTGCGCCCGAGACTCCAAATCCCACTGATCTGCTCCATGAGCCG 302

Qy 301 GlnAlaAlaHisHisLeuProHisHisHisHisHisHisHisHisHisHisHisHisHis 320

Db 303 GAAGCATTCACATCCACACAG 362

Qy 321 LeuAspThrProIleAlaLysValSerGlnGlnGlnArgLeuArgGlyThrGlnAsp 340

Db 363 CTTGACACCCCAATCGCCAGAGTCTCAGAGCTCAGACAGAGCTCCGGGACACCCAGAGAC 422

Qy 341 GlySerTrpHisPheValArgSerProLysValGlnGlyLysSerValGlyValGlyHis 360

Db 423 GGAAGCAAGACTTGTGAGGTCTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 482

Qy 361 ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380

Db 483 GTGGCCAGAGGGGACAGAAACAGACCCCTCTGGAGCCGCGCATTCCTGCGGTCTCC 542

Qy 381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHlyHly 400

Db 543 TCCGCCACCTGGCTGCTGAGCCGCGGCTCTCTCCCTTACAGCCCTCGGAGCAAG 602

Qy 401 LysHisLysHisHisArgAlaLysGlnSerGlnGlnGlyCyAArgGlyLeuGlnAlaProLeu 420

Db 603 AAGCACAAGACGAGCCAG 662

Qy 421 AlaSerGlyGlyProValLeuGlyArgGlnHisLeuArgGlnLeuProAlaLeuVal--- 439

Db 663 GCTCAGGTGGCTCTGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 722

Qy 440 ValTyrgLuserGlnAlaGlyGlnProVal---GlnArgHisGlnHisHisHisHisHis 458

Db 723 GTATGAGAACCAAGCCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 782

Qy 458 Gln 478

Db 783 TGA 786

RESULT 14

LOCUS AY412099 885 bp DNA linear GSS 16-DEC-2003

DEFINITION Pan troglodytes NKD1 gene, VIRUAL TRANSCRIPT, partial sequence, genomic survey sequence.

ACCESSION AY412099

VERSION AY412099.1 GI:39768064

KEYWORDS GSS.

SOURCE Pan troglodytes (chimpanzee)

ORGANISM Pan troglodytes

REFERENCE	AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.	
1 (baaes 1 to 885)	
Clark,A.G., Gnanowsk,i,S., Nielson,R., Thomas,P., Kejarial,A., Todd,M.A., Tannenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Perrier,a,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.J., Adams,M.D. and Cargill,M.	
ADAMS,M.D. and CARGILL,M. Direct Submission Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA	
This sequence was made by sequencing genomic exons and ordering them based on alignment	
FEATURES	
source	Location/Qualifiers
gene	1..885 /organism="Pan troglodytes" /mol_type="genomic DNA" /db_xref="taxon:9598" <1...>885 /gene="NKD1" /locus_tag="HCW4444"
ORIGIN	
Alignment Scores:	
Pred. No. :	3,06e-90 Length: 885
Score:	1240.00 Matches: 247
Percent Similarity:	84.41% Conservative: 2
Best Local Similarity:	83.73% Mismatches: 46
Query Match:	49.62% Indels: 0
DB:	Gaps: 0
US-93-993-966-7 (1-470) x AY412099 (1-885)	
OY	65 GIULEUValGlyAspValLeuArgAapThrLeuSerGIuGIuGIuAapPheArg 84
Dd	1 GAGCTCGTGGGGGACGCTGTGGAGACACGCCTCAGCGAGGAAGAAGGACGACTTCGG 60
OY	85 LeuGIuValAlaLeuProProGIuLysTrpAspGIyLeuGIySerGIyAapGIyLys 104
Dd	61 CTGGAATGTGCCCTTGCTCTGTAAGACTGACGGGCTGGGCAAGCGAATGAAAGAAG 120
OY	105 MetGIuArgValSerGIuProCySProGIySetLyEllysGIuLeuLysPheGIuGIuLeu 124
Dd	121 ATGGAGAGAGTAGTACGACCCCTGCCAGGCTCCAAGAAGCAGCTGAAGTTGAAGAGCTC 180
OY	125 GInCYbAspValSerMetGIuGIuAapSerArgInGIuITrPhethrLeuTyrrAsp 144
Dd	181 CAGTGTGACGTGTCCCTGGAGAGGACAGCCGCGACAGAGTGGACCTTCACCTGTATGAC 240
OY	145 PheaAspAsnAsnGIyLysValThrArgGIuAapGlLeThrSerLeuLeuHsthrIleTy 164
Dd	241 TTtTGACACAACGGCACAGGTACCAGAGAGAACATCACAGCTTGCTGCACACCATTAT 300
OY	165 GIuValAlaAspSerSerValAsnHisSerProThrSerSerLyMetLeuArGVallys 184
Dd	301 GAGGTGTGTGACTCCCTGTGTCAACCACTCCCCAACATCCAGCAAAGATGCTGGCGGTAAG 360
OY	185 LeuthrValAlaProAapGIySerGIuSerLysArgSerValLeuValAsnGINalAap 204
Dd	361 CTCACCGTGGCCCCCATGTGACAGCCAGAGCAAGAAGAGCGTCCCTTGCAATCAGGCTGAC 420
OY	205 LeuGIuSerAlaArgProArgAlaGIuTrpLysProThrGIuAspLeuAArgserTPGu 224
Dd	421 CTGCAGAGCGCAAGGCCCCGAGCAGAAACCAACCCCATGTAGGAGCTGGCGAGCTGGAG 480

[illegible]

ORIGIN

Alignment Scores:

Pred. No.:	3,06e-88	Length:	966
Score:	1216.00	Matches:	233
Percent Similarity:	97.91%	Conservative:	1
Best Local Similarity:	97.49%	Mismatches:	4
Query Match:	48.66%	Indels:	1
DB:	5	Gaps:	0

US-09-993-966-7 (1-470) x BQ649813 (1-966)

```

QY      201  ASNGIALASPLeuGInSerAlaArgProArgAlaGluThrLysProThrGluASPLeu 220
DB      2    CATCAGCTGACCTGCGAGAGCGGAGGCCCGAGACAGACCAAGCCACTGAGGACCTG 61
QY      221  ArgSerTrpGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
DB      62  CGGAGCTGGGAGAGAGAGAGCGAGCCCGCTCAGGTTCCAGGGTGAACGCCGCTGGAG 121
QY      241  GlnSerGlyCybTyHisHisCysValAspGluAsnIleGluArgArgAsnHisTyrLeu 260
DB      122  CAGCTGCTGCTGCTACCACTTGCTGATGAGAAACATCGAGAGAGAAACCACTACTTA 181
QY      261  ASPLeuAlaGlyIleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValAla 280
DB      182  GATCTCGCCGGGATGAGAAACTACACGTCCTCAATTGGGCTGGCTCCCTCCGCTGGGCC 241
QY      281  GlnLysSerGluLeuProArgThrSerAspProThrArgSerArgSerHisGluPro 300
DB      242  CAGAGTCAGAGACTGCCCCCGGCACTCCAAATCCCACTGATCTCGCTCCCATGAGCCG 301
QY      301  GluAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPhe 320
DB      302  GAAGCCATCCACATCCACACCGAAAGCCCAAGCGGTGACCCGCTCCCTCCACTTC 361
QY      321  LeuAspThrProIleAlaLysValSerGluLeuGlnGlnArgLeuArgGlyThrGlnAsp 340
DB      362  CTGGACACCCCAATGCCAAGGTCTCAGAGCTCCAGCAAGGCTCCGGGGCACCAGGAC 421
QY      341  GlySerLysHisPheValArgSerProLysPheAlaGlnGlyLysSerValGlyValGlyHis 360
DB      422  GGGAGCAAGCACTTGTGAGGTGCCCAAGGCCCAAGGCAAGAGTGTGGGTGGGCAC 481
QY      361  ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
DB      482  GTGGCCAGAGGGGCAAGAAACAAGCCCCCTGTGGACCCGCCATCCTGCGGTGTCCCC 541
QY      381  SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
DB      542  TCCGCCCACTGGCTGCAAGCCCGGGGCTCTCTCCCTCCCTAGCCCCCTCGGGCAACG 601
QY      401  LysHisLysHisValArgAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
DB      602  AAGCAACAAGCAACGAGCCAGAGAGAGCAAGCAAGGCTGCCGGGGCTTCAGGCACTG 661
QY      421  AlaSerGlyGlyProValLeuGlyArgGlnHisLeu-ArgGluLeuProAlaLeu 438
DB      662  GCCTCAGGTGGCTGTCTGGGGCGGAGACNACTGCGGAGGCTTGCGCGGCTTG 716

```

Search completed: December 30, 2004, 07:54:26
 Job time : 5660 secs

THIS PAGE IS BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 30, 2004, 03:48:50 ; Search time 5778 Seconds

(without alignments)
459.446 Million cell updates/sec

Title: US-09-993-966-7

Perfect score: 2499

Sequence: 1 MGKLSKPAVCKRRSPG.....RHHHHHHHHHHFYQT 470

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4176236 seqs, 2824127955 residues

Total number of hits satisfying chosen parameters: 8352472

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model /US09993966/runat 29122004.130759.9062/app query.fasta_1.647
-O=/cgn2.1/USPTO.spool.p/US09993966 -DEV=xip
-DB=Published Applications NA -QFMT=fastaj -SUFFIX=rnpb -MINMATCH=0.1
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human0.csl -LIST=45 -DOCALIGN=200 -THR SCORE=PCT -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pico -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09993966 @CGN 1.1 912 @runat 29122004.130759.9062
-NCPUS=6 -ICPU=3 -NO MMAP -LARGEOUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FARPOP=6 -FARPEXT=7 -YCAPOP=10 -YCAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2.6/prodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2.6/prodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2.6/prodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2.6/prodata/1/pubpna/US07_NEW_PUB.seq:*
5: /cgn2.6/prodata/1/pubpna/US06_PUBCOMB.seq:*
6: /cgn2.6/prodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2.6/prodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2.6/prodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2.6/prodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2.6/prodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2.6/prodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2.6/prodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2.6/prodata/1/pubpna/US10_PUBCOMB.seq:*
14: /cgn2.6/prodata/1/pubpna/US10_PUBCOMB.seq:*
15: /cgn2.6/prodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2.6/prodata/1/pubpna/US10_PUBCOMB.seq:*
17: /cgn2.6/prodata/1/pubpna/US10_PUBCOMB.seq:*
18: /cgn2.6/prodata/1/pubpna/US10_NEW_PUB.seq:*
19: /cgn2.6/prodata/1/pubpna/US11_NEW_PUB.seq:*
20: /cgn2.6/prodata/1/pubpna/US60_NEW_PUB.seq:*
21: /cgn2.6/prodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2499	100.0	1448	13 US-10-087-192-179	Sequence 179, App
2	2499	100.0	1786	10 US-09-993-966-1	Sequence 1, Appl1
3	2499	100.0	1859	10 US-09-993-966-5	Sequence 5, Appl1
4	2189.5	87.6	1416	10 US-09-993-966-2	Sequence 2, Appl1
5	2171.5	86.9	1401	10 US-09-993-966-6	Sequence 6, Appl1
6	2167.5	86.7	1401	9 US-09-730-989-1	Sequence 1, Appl1
7	2085.5	83.5	2807	13 US-10-087-192-176	Sequence 176, App
8	1191.5	47.7	106315	13 US-10-087-192-178	Sequence 178, App
9	1076	43.1	98474	13 US-10-087-192-175	Sequence 175, App
10	1067	42.7	2379	10 US-09-764-891-5992	Sequence 5902, App
11	1067	42.7	2379	10 US-09-764-891-5994	Sequence 5904, App
12	1053	42.1	500	10 US-09-993-966-21	Sequence 21, Appl1
13	831.5	33.3	1966	9 US-09-764-868-152	Sequence 152, Appl1
14	786	31.5	1959	18 US-10-723-860-5220	Sequence 5220, App
15	739	29.6	427	10 US-09-918-995-8686	Sequence 8686, App
16	378.5	15.1	1837	9 US-09-822-849A-83	Sequence 83, Appl1
17	274	11.0	73038	17 US-10-322-696-37	Sequence 37, Appl1
18	271	10.8	639	15 US-10-243-552-231	Sequence 31, App
19	259	10.4	901	13 US-10-027-632-162722	Sequence 162722, Sequence 162722,
20	259	10.4	901	13 US-10-027-632-162722	Sequence 126, App
21	246	9.8	1950	13 US-10-016-157A-7990	Sequence 7990, App
22	239	9.6	598	15 US-10-029-386-7990	Sequence 18, Appl1
23	238	9.5	148	10 US-09-993-966-18	Sequence 20, Appl1
24	237	9.5	128	10 US-09-993-966-20	Sequence 21690, A
25	233	9.3	133	15 US-10-029-386-21690	Sequence 14, Appl1
26	233	9.3	134	10 US-09-993-966-14	Sequence 16, Appl1
27	186	7.4	107	10 US-09-993-966-16	Sequence 17, Appl1
28	175	7.0	96	10 US-09-993-966-17	Sequence 17, Appl1
29	172	6.9	6749	18 US-10-723-860-6868	Sequence 6868, App
30	161.5	6.5	554	15 US-10-029-386-7615	Sequence 7615, App
31	161	6.4	2481	18 US-10-425-115-114040	Sequence 114040,
32	156.5	6.3	1633	15 US-10-017-161-1819	Sequence 1819, App
33	156.5	6.3	1633	15 US-10-292-798-1475	Sequence 1475, App
34	156.5	6.3	155073	13 US-10-026-188-6	Sequence 6, Appl1
35	155	6.2	84428	16 US-10-229-148B-1	Sequence 1, Appl1
36	153	6.1	744802	9 US-10-892-798-1369	Sequence 1369, App
37	152	6.1	2433	15 US-09-944-807-3	Sequence 3, Appl1
38	152	6.1	2433	15 US-10-290-198-2	Sequence 2, Appl1
39	152	6.1	2433	15 US-10-172-118-1068	Sequence 1068, App
40	152	6.1	2433	16 US-10-342-887-1068	Sequence 1068, App
41	152	6.1	2433	18 US-10-874-015-3	Sequence 3, Appl1
42	151.5	6.1	2078	15 US-10-307-047-35	Sequence 35, Appl1
43	151	6.0	3369	18 US-10-786-720-7	Sequence 8, Appl1
44	151	6.0	3477	18 US-10-786-720-7	Sequence 7, Appl1
45	150.5	6.0	2978	14 US-10-116-802-184	Sequence 184, App

ALIGNMENTS

RESULT 1
US-10-087-192-179
Sequence 179, Application US/10087192
Publication NO. US20020182586A1
GENERAL INFORMATION:
APPLICANT: Morris, David W.
APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
FILE REFERENCE: 529452000122
CURRENT APPLICATION NUMBER: US/10/087,192
CURRENT FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 2059
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 179
LENGTH: 1448
TYPE: DNA

ORGANISM: Homo sapiens
US-10-087-192-179

Alignment Scores:

Pred. No.:	3.52e-227	Length:	1448
Score:	2499.00	Matches:	470
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	13	Gaps:	0

US-09-993-966-7 (1-470) x US-10-087-192-179 (1-1448)

```

QY 1 MetGlylyleuHisSerlySProAlaAlaValCyelybArGaRgJuserProGluGly 20
Db ATGGGGAAACCTTCACTCCAAAGCCGGCCGGTGTGCAAGCGCAAGAGCCCGAAAGGT 98
QY 21 AspSerPheAlaValSerAlaAlaTPrAlaRghlySgJlyIleGluJUTPrIleGlyArg 40
Db GACAGCTTCCCGCGAGCGCTGCTGGGCTCGAAAGGCATCGAGAGTGGATCGGAGA 158
QY 41 GluArgCyAProGlyGlyValSerGlyProArgGluLeuArgLeuAlaGlyThrIleGly 60
Db CAGGCTGCCCGGGCGGTGTCTCGGAGACCCGACAGCTGCGGTTGGCGGGACCATAGGC 218
QY 61 ArgSerThrArgGluLeuValAlaAspValLeuArgPheThrLeuSerGluGluGlu 80
Db 219 CGAAGCACCCGGGAGCTGTGGGCGAGCTGTGAGACACAGCTGCACGAGAGAGAG 278
QY 81 AspAspPheArgLeuGluValAlaLeuProProGluThrArgPGLyleuGlySerGly 100
Db 279 GACGACTTGGGCTGGAGAGTGGCCCTGCTCTGAGAAAGCTAGCGGGCTGGGAGCGA 338
QY 101 AspGluValylsMetGluArgValSerGluProCyAProGlySerlySgJlyLeuLys 120
Db 339 GATGAGAAAGATGAGAGAGTGAAGCAACCTGCCAGGCTCCAAGAGCAGCTGAAG 398
QY 121 PheGluGluLeuGlnCyAProValSerMetGluGluLeuSerArgGlnGluTPrPhe 140
Db 399 TTGGAAGAGCTCCAGTCCGAGCTGTCATGAGAGAGAGAGCCGGAGAGTGAACCTTC 458
QY 141 ThrLeuTyRAspPheAspAsnGlylyValValThrArgIleuAspIleThrSerLeu 160
Db 459 ACCCTGATGACTTTGACACACAGGCAAGTCAACCGAGAGGACATCACCAGCTTGTG 518
QY 161 HisThrIleTyrgluValAlaAspSerSerValAsnHisSerProThrSerSerlyMet 180
Db 519 CACACCATATGAGGTGTGACTCTCTGTCAACCACTCCCAACATCCAGCAAGATG 578
QY 181 LeuArgValylsLeuThrValAlaProAspGlySerGlnSerlybArGserValLeuVal 200
Db 579 CTGGGGTAAAGCTCACCGTGGCCCGATGGCCAGCAAGAGAGAGAGGCTCTTGTG 638
QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrlySgJlyProThrGluAspLeu 220
Db 639 AATAGGCTGACTGCGAGGCGCAAGGCCCGAGCAGAGACCAAGCCCACTGAAGACCTG 698
QY 221 ArgSerTrpGluValylsGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
Db 699 CGGAGCTGGGAGAGAGAGCGAGCCCGCTCAAGTTCCAGGTTGACAGCGCTGAGAG 758
QY 241 GluSerGlyCybTyRHisGlyValAspGluAsnIleGluArgArgAsnHisTyRLeu 260
Db 759 CAGCTGTGCTGCTACCACTATGCTGATGAGACATCGAGAGAGAGAGAGAGAGAGAGAG 818
QY 261 AspLeuAlaGlyIleGluAsnTyRThrSerGlnPheGlyProGlySerProSerValAla 280
Db 819 GATCTCCCGGAGATGAAACTACAGTCCCAATTTGGGCTGCTCCCTCCGTGAGCC 878
QY 281 GluLysSerGluLeuProArgThrSerAsnProThrArgSerArgSerHisGluPro 300
Db 879 CAGAGTCAAGAGAGTGGCCCGGCACTCCCAATCCCAATGATTCGCTCCCAAGAGCCG 938

```

```

QY 301 GluAlaIleHisIleProHisArgLySProGlnGlyValAspProAlaSerPheHisPhe 320
Db 939 GAGCCATCCATCCATCCACACCGAAAGCCCGAGGCTGAGACCGGCTCTTCACTTC 998
QY 321 LeuAspThrProIleAlaValylsSerGluLeuGlnGluArgLeuArgJlyThrGlnAsp 340
Db 999 CTTGACACCCCAATCCGCAAGGTCTCAGAGCTCCAGCAAGCGCTCCGGGCAACCAAGAG 1058
QY 341 GlySerlybHisPheValaArgSerProLybAlaGlnGlylySgSerValGlyValyls 360
Db 1059 GGGAGCAAGCACTTGTGAGGTCCCGCAAGGCCCGCAAGAGTGTGGGTGTGGGCAC 1118
QY 361 ValAlaArgGlyAlaArgAsnLySProProLeuGlyProAlaIleProAlaValSerPro 380
Db 1119 GTGGCCAGAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1178
QY 381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
Db 1179 TCGGCCACCTGGGTGCAAGCCCGGCTCTCCCTCCCTAGCCCCCTCGGGCACAG 1238
QY 401 LysHisLybHisArgAlaValylsSerGlnGlnGlyCybArgGlyLeuGlnAlaProLeu 420
Db 1239 AAGCACAGAGCAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1298
QY 421 AlaSerGlyGlyProValaArgGlyArgGluHisLeuArgGluLeuProAlaLeuVal 440
Db 1299 GCCTCAGGTGACCTGTCTGTGGGCGGAGACCTGCGGAGCTGCCCGCTGTGTGTG 1358
QY 441 TyrgluSerGlnAlaGlyGlnProValaGlnArgHisGluHisHisHisHisGluHis 460
Db 1359 TATGAGAGCCAGGCGGGGCGAGCGGTCCAGAGACATGAGACCAACCAACATGACAT 1418
QY 461 HisHisIleTyRHisPheTyRGINThr 470
Db 1419 CACCACTTACCACTTCTACAGACA 1448

```

RESULT 2

```

US-09-993-966-1
: Sequence 1, Application US/09993966
: Publication No. US20030186232A1
: GENERAL INFORMATION:
: APPLICANT: ROHAN, MICHAEL
: TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NKD PROTEIN,
: FILE REFERENCE: 014024/0280733
: CURRENT APPLICATION NUMBER: US/09/993,966
: PRIOR APPLICATION NUMBER: 60/252,884
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/291,109
: PRIOR FILING DATE: 2001-05-16
: PRIOR APPLICATION NUMBER: 60/325,571
: PRIOR FILING DATE: 2001-10-01
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1
: LENGTH: 1786
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-993-966-1

```

Alignment Scores:

Pred. No.:	4.62e-227	Length:	1786
Score:	2499.00	Matches:	470
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

US-09-993-966-7 (1-470) x US-09-993-966-1 (1-1786)

```

QY 1 MetGlylyleuHisSerlySProAlaAlaValCyelybArGaRgJuserProGluGly 20

```

```

Db      159 ATGGGAAACTTCACTCCAAAGCCGCCGTGTGCAAGCGCAAGAGAGAGCCGGAAGCT 218
QY      21 ApsSerPheAlaValSerAlaAlaATPAlaArgLysGlyIleGluTTPiIleGlyArg 40
Db      219 GACAGCTTCGCGCTGAGCGCTGCGCTGGGCTCGGAAGGCGATCGAGAGTGGATCGGAGA 278
QY      41 GlnArgCysProGlyValValSerGlyProArgGlnLeuArgLysIleGlyThrIleGly 60
Db      279 CAGCGCTGCGCGCGGTGTCTCGGAGCCCGACAGCTGGCGTTGGGGGCAACATAGGC 338
QY      61 ArgSerThrArgGlyLeuValGlyAapValLeuArgPheThrLeuSerGlnGluGlu 80
Db      339 CGAAGCAACCCGGAGCTCTGGCGGAGCTGTGAGAGACAGCTCAGCGAGGAAGAGAG 398
QY      81 AapPhePheArgLeuGluValAlaLeuProProGlyLysThrArgGlyLeuGlySerGly 100
Db      399 GACGACTTTCGCTGGAAAGTGCCCTGCTCTGAGAAAGACTGACGGGCTGGGACGGA 458
QY      101 AapGlyLysLeuMetGlyLysValSerGluProCysProGlySerLysLysGlnLeuLys 120
Db      459 GATGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 518
QY      121 PheGlyLysLeuGlnCysAapValSerMetGlyLysAapSerArgGlnGluTTPiIlePhe 140
Db      519 TTGGAAGAGCTCAGTCCGACGTCTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGCTTC 578
QY      141 ThrLeuTyrAapPheAapAaAaAnglyLysValThrArgLysAapIleThrSerLeuLeu 160
Db      579 ACCCTGATGACTTTGACAAACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 638
QY      161 HisThrIleTyrLysValValAapSerSerValAaHisSerProThrSerSerLysMet 180
Db      639 CACACACATCATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 698
QY      181 LeuArgValLysLeuThrValAlaProAapGlySerGlnSerLysAapSerValLeuVal 200
Db      699 CTGCGGAGTAAAGCTCACCGTGGCCCGGATGGCAGCGAGAGAGAGAGAGAGAGAGAGAG 758
QY      201 AapGlnAlaAapLeuGlnSerAlaArgProArgAlaGluThrLysProThrGlyAapLeu 220
Db      759 AATCAGAGCTGACTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 818
QY      221 ArgSerTyrGlyLysLysGlnArgAlaProLeuArgPheGlnGlyAapSerArgLeuGlu 240
Db      819 CGAGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 878
QY      241 GlnSerGlyCysTyrTrpHisSerValAapGlyAaHisIleGluArgAaAaHisTyrLeu 260
Db      879 CAGTCTGGCTGCTACCACTTGGTGTGATGAGACATCGAGAGAGAGAGAGAGAGAGAGAG 938
QY      261 AapLeuAlaGlyIleGluAaTyrTrpSerGlnPheGlyProGlySerProSerValAla 280
Db      939 GATCTGCGCGGAGTAAAGAACTACAGTCCCAATTTGGGCTGGGCTCCCTCCGCTGGCC 998
QY      281 GlnLysSerGlyLeuProProArgThrSerAapProThrArgSerArgSerHisIleGluPro 300
Db      999 CAGAAATCGAAATCGCCCGCCGACCTCCAAATCCCACTCGATCTCCCTCCCATGAGCCG 1058
QY      301 GlnAlaIleHisIleProHisArgLysProGlyValAapProAlaSerPheHisPhe 320
Db      1059 GAAAGCATTCACATCCCAACCGAAAGCCCAAGGCGTGAACCGGCGCTCTTCCACTTC 1118
QY      321 LeuAapThrProIleAlaLysValSerGlyLeuGlnGlnArgLeuArgGlyThrGlnAap 340
Db      1119 CTTGACACCCCAATCGCCAGAGCTCAAGAGCTCAGAGACGGGCTCGGGGCAACCAAGAGC 1178
QY      341 GlySerLysHisPheValArgSerProLysValGlnGlyLysSerValGlyValGlyHis 360
Db      1179 GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1238
QY      361 ValAlaArgGlyValAaArgAaLysProProLeuGlyProAlaIleProAlaLysPro 380
Db      1239 GTGGCCAGAGGGGCAAGAAACAAAGCCCTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1298

```

```

QY      381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
Db      1299 TCCGCCCACTTGGCTGCAAGCCCGGCGCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1358
QY      401 LysHisIleHisPheArgAlaLysGlnSerGlnGlyCysAapArgGlyLeuGlnAlaProLeu 420
Db      1359 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1418
QY      421 AlaSerGlyLysProValLeuGlyArgGlnHisLeuArgGlyLeuProAlaLeuVal 440
Db      1419 GCCTCAGAGTGGCCCTGTCTGGGGCGGAGACACTGCGGAGAGTGGCCGCTGTGTGTGTG 1478
QY      441 TyrGlnSerGlnAlaGlyLysProValGlnArgHisIleGlnHisPheHisIleGlnHis 460
Db      1479 TATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1538
QY      461 HisHisIleTyrHisPheTyrGlnThr 470
Db      1539 CACCACTTACCAACCACTTACCAAGACA 1568

RESULT 3
US-09-993-966-5
; Sequence 5, Application US/09993966
; Publication NO. US20030186232A1
; GENERAL INFORMATION:
; APPLICANT: ROHAN, MICHAEL
; TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NKD PROTEIN,
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF
; FILE REFERENCE: 014024/0280733
; CURRENT APPLICATION NUMBER: US/09/993,966
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 60/252,884
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/291,109
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/325,571
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1859
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-993-966-5

Alignment Scores:
Pred. No.: 4 866-227 Length: 1859
Score: 2499.00 Matches: 470
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-09-993-966-7 (1-470) x US-09-993-966-5 (1-1859)
QY      1 MetGlyLysLeuHisSerLysProAlaAlaValCysLysArgArgGlnSerProGlnGly 20
Db      224 ATGGGAAACTTCACTCCAAAGCCGCCGTGTGCAAGCGCAAGAGAGAGAGAGAGAGAGAG 283
QY      21 ApsSerPheAlaValSerAlaAlaATPAlaArgLysGlyIleGluTTPiIleGlyArg 40
Db      284 GACAGCTTCGCGCTGAGCGCTGCGCTGGGCTCGGAAGGCGATCGAGAGTGGATCGGAGA 343
QY      41 GlnArgCysProGlyValValSerGlyProArgGlnLeuArgLysIleGlyThrIleGly 60
Db      344 CAGCGCTGCGCGCGGTGTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 403
QY      61 ArgSerThrArgGlyLeuValGlyAapValLeuArgAapThrLeuSerGlnGluGlu 80
Db      404 CGAAGCAACCCGGAGCTCTGGGCGAGAGTGTGAGAGACAGCTCAGCGAGAGAGAGAGAGAG 463
QY      81 AapPhePheArgLeuGluValAlaLeuProProGlyLysThrArgGlyLeuGlySerGly 100

```

```
Db 464 GAGACCTTTGGGTGGAGAGTGGCTCTCTCTGAGAGAGCTAGAGGGGCTGGGCGAGCGGA 523
Qy 101 AaPGLuLySerMetGluValSerGluProCysProGlySerIlySgInLeuLys 120
Db 524 GATGAGAGAGAGATGAGAGATGAGAGAGCTGCGCCAGGCTCCAGAGAGAGAGCTGAG 583
Qy 121 PheGluGluLeuInCysAaPValSerMetGluGluAaSerArgGlnGluTTPThPhe 140
Db 584 TTGGAAGAGCTCAGTGCAGCTGTCTCATGAGAGAGAGAGCGGCGAGGTGAGCTTTC 643
Qy 141 ThLeuTyAaPheAaPaaAnGlyValThrArgGluAaSplIethrSerLeuLeu 160
Db 644 ACCCTGTATACCTTGACACCAAGGAGAGGTACCCGAGAGAGACATCACAGCTTCTG 703
Qy 161 HisThrIleTyGluValValAaPserSerValAaAaHisSerProThrSerSerLysMet 180
Db 704 CACACCATATATAGGTGTGAGCTCTGTCTGCAACCACTCCCAACATCCAGAGATG 763
Qy 181 LeuArgValIlySerThrValAlaProAaPGLySerGlnSerIlyAaAaSerValLeuVal 200
Db 764 CTGCGGGTAAAGCTCACCGTGGCCCCGATGGCAGCCAGAGCAAGAGAGAGCTTCTGTC 823
Qy 201 AaArgIAlaAaPLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAaPLeu 220
Db 824 AATCAGGCTGACCTGCAGAGCGCAGAGGCCCGAGCAGAGACCAAGCCCATGAGAGACTG 883
Qy 221 ArgSerITPGLuLySgInArgAlaProLeuArgPheGlnGlyAaPserArgLeuGlu 240
Db 884 CGAGAGCTGGAGAGAGAGAGAGAGCGCCGCTCAGGTTCAGAGGTGACAGCCGCTGAG 943
Qy 241 GlnSerGlyCySerTyHisSgysValAaPGLuAaSplIethrArgAaAaHisTyLeu 260
Db 944 CAGTCTGCTGTACCAACATGCGTAGATGAGAAACATCAGAGAGAGAAACCACTACTTA 1003
Qy 261 AaPLeuAlaGlyIleGluAaAaTyThrSerGlnPheGlyProGlySerProSerValAla 280
Db 1004 GATCTGCGCGGAGTAAAGAACTACAGCTCCCAATTTGGGCTGGCTCCCTTCCGTGGCC 1063
Qy 281 GlnLysSerGluLeuProProArgThrSerAaPProThrArgSerArgSerHisGluPro 300
Db 1064 CAGAAAGTCAAGAACTGCCCCCGCACCTCCAAATCCCATGCTGCTCCCATGAGACCG 1123
Qy 301 GlnAlaIleHisIleProHisArgLysProGlnGlyValAaPProAlaSerPheHisPhe 320
Db 1124 GAAGCCATTCACATCCACACCGAAAGCCCAAGGCGGTGAGCCGCGCTCTTCCACTTC 1183
Qy 321 LeuAaPThrProIleAlaLysValSerGluLeuGlnGlnArgLeuArgGlyThrGlnAaP 340
Db 1184 CTTGACACCCCAATCCGCAAGGTTCAGAGCTCCAGCAGAGCGCTCCGGGCGACCCAGAG 1243
Qy 341 GlySerIlyHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValGlyHis 360
Db 1244 GGGAGCAGAGCACTTGTGAGGTCCCCCAAGGCCCAAGGAGTGTGGGTGTGGGCAC 1303
Qy 361 ValAlaArgGlyAlaArgAaAaLysProProLeuGlyProAlaIleProAlaValSerPro 380
Db 1304 GTGGCCAGAGGGGCAAGAAACAAAGCCCCCTGTGGAGCCCGCATCTCCGTGCTCCCC 1363
Qy 381 SerAlaHisLeuAlaAaSerProAlaLeuAaPProSerLeuAlaProLeuGlyHisLys 400
Db 1364 TCCGCCCACTGGGTGCGCAGCCCGGCTCTCTCCCTCTAGGCCCTCCGCGGCAAG 1423
Qy 401 LysHisIlyHisAaGAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
Db 1424 AAGCACAAGACACGAGCAAGAGAGAGAGAGGAGTGGCGGGGCTGACAGGCAACACTG 1483
Qy 421 AlAaSerIlyGlyProValLeuGlyArgGluHisIleuArgGluLeuProAlaLeuValVal 440
Db 1484 GCTCTAGGTGACCTGTCTGTGGGGGAGAGACCTTGGGAGCTGCCCGCTTGTGTG 1543
Qy 441 TyGlnSerGlnAlaGlyGlnProValGlnArgHisGlnHisSgHisSgHisGlnHis 460
```

```
Db 1544 TATGAGAGCCAGGCGGCGAGCCGCTCCAGAGATGAGAGCAGCAGCAACATGACAT 1603
Qy 461 HisHisIleTyThrHisHisPheTyArgIleThr 470
Db 1604 CACCACCATTAACCAACCATTTACAGAGACA 1633

RESULT 4
US-09-993-966-2
; Sequence 2, Application US/09993966
; Publication No. US20030186232A1
; GENERAL INFORMATION:
; APPLICANT: ROHAN, MICHAEL
; TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NR2 PROTEIN,
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF
; FILE REFERENCE: 014024/0280733
; CURRENT APPLICATION NUMBER: US/09/993, 966
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 60/252, 884
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/291, 109
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/325, 571
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Mus sp.
US-09-993-966-2

Alignment Scores:
Pred. No.: 7 79e-198 Length: 1416
Score: 2189.50 Matches: 409
Percent Similarity: 91.06% Conservative: 19
Best Local Similarity: 87.02% Mismatches: 41
Query Match: 87.62% Indels: 1
DB: Gaps: 1

US-09-993-966-7 (1-470) x US-09-993-966-2 (1-1416)
Qy 1 MetGlyLysLeuHisSerIlySerProAlaValCysLysArgArgGluSerProGlnGly 20
Db 1 ATGGGGAAATTCACCTGAGAGCCGGCGCGTGTGCAAGCGCAGAGAGAGAGAGAGT 60
Qy 21 AaPserPheAlaValSerAlaAlaATPAlaArgLysGlyIleGluGluTTPIleGlyArg 40
Db 61 GACAGCTTGTCTGTAAAGCGCTGCTGGCAAGAAAGCAATCGAGAGGTGATCGGAGG 120
Qy 41 GlnArgCysProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGly 60
Db 121 CAGGCTGTCTCAGGACAGCGCTCTCGAGAGCCCGTCAAGTGAATTTGGCAGGCACTGTGGT 180
Qy 61 ArgSerThrArgGluLeuValGlyAaPValLeuArgAaPThrIleuSerGlnGluGlu 80
Db 181 CGAGGCACTGGGGAAGTGTGGGTGACACTTCTAGAGAGGCTCTCGGTGAGAGAGAG 240
Qy 81 AaPserPheArgLeuGluValAlaLeuProProGlnLysThrAaPGLyLeuGlySerGly 100
Db 241 GAGGACTTCCCCCTAGAAAGTGGCCCTGCCCTGAGAAAGATCAACAGCTAAGTATGGA 300
Qy 101 AaPGLuLySerMetGluValSerGluProCysProGlySerIlySgInLeuLys 120
Db 301 GATGAGAGAGAGATGAGAGACTGAGAGAGCTGGCCAGGCTCCAGAGAGAGAGCTAG 360
Qy 121 PheGluGluLeuInCysAaPValSerMetGluGluAaSerArgGlnGluTTPThPhe 140
Db 361 TTGGAAGAGCTACAGTGTATGTCTGTGTGAGAGAGAGAGCGGCAAGAGTGAAGTTC 420
Qy 141 ThLeuTyAaPheAaPaaAnGlyValThrArgGluAaSplIethrSerLeuLeu 160
Db 421 ACTTATATACCTTGACACCAATGGCAAGTGAACCGGTGAGGACATTAACAGCTTCTG 480
```

```

Oy 161 HistHrIleTyrGluValValaAspSerSerValaAsnHisSerProThrSerSerIleuMet 180
Db 481 CATACCATCTATGAAAGTGGTTGACTCTCTGTGAACATTTCCCACTCAAGAGCA 540
Oy 181 LeuArgValIleuLeuThrValaIaProAspIleuSerIleuValaIleuVal 200
Db 541 CTCGGGGTGAAGCTCACCGTGGCTCTGACGGGAGCCAGATGAAGAGAGTCTTTTC 600
Oy 201 AsnGlnIleuAspLeuGlnSerIleuValaIaArgProArgIleuValaIleuVal 220
Db 601 AACCATACCATCTGACAGACCAAGAGCCCGAGCAGACCAACCCCGTGAAGAGCTG 660
Oy 221 ArgSerTrpGluValValaIaArgProArgIleuValaIleuValaIleuVal 240
Db 661 CTTGGCTGGAGAAAGAGCAGAGCCCACTCAGTTCCAGGTTGACAGCCTGAG 720
Oy 241 GlnSerGlyCysTrpHisIleuValaIaAspIleuValaIleuValaIleuVal 260
Db 721 CAGCAGACTGCTACCACTTGGTGGATGAGAACTTGAAGAGAAACCACTACCTA 780
Oy 261 AspLeuValaIleuValaIleuValaIleuValaIleuValaIleuValaIleu 280
Db 781 GACCTGGCGGAGTAGAGAACTACACGCTCAGTTTGAACCGGAGATCCCTTCGGTGGC 840
Oy 281 GlnIleuSerGluLeuProArgTrpSerAsnProThrArgSerArgSerHisGluPro 300
Db 841 CAGAAAGTCAGAGCTGCCCCCTCGAATCTCAACCCCTGCTCTCCCTCCCAAGACCA 900
Oy 301 GlnValaIleuHisIleuProHisIleuValaIleuValaIleuValaIleuVal 320
Db 901 GAAGCTGCCACATCCCAACCGAGGAGCCCAAGGTGAGACCAAGGCTCTTCCACTC 960
Oy 321 LeuAspThrProIleuValaIleuValaIleuValaIleuValaIleuValaIleu 340
Db 961 CTGGACACCCCATTTGGCAAGCATCAGACTCCAGCAACGAGCTCCGGGCACTCAGAGT 1020
Oy 341 GlnSerTrpHisIleuValaIleuValaIleuValaIleuValaIleuValaIleu 360
Db 1021 GGGAGCAGAGCACTTGTGAAGTCCCCCAAGGAGCCCAAGCAAGATGGATGGGCAC 1080
Oy 361 ValaIleuValaIleuValaIleuValaIleuValaIleuValaIleuValaIleu 380
Db 1081 GGGGCGCAGAGTGCAGAGAGCAAGCCCTCAGTGTACCAACCCCTACTGTTCCCTCC 1140
Oy 381 SerAlaHisIleuValaIleuValaIleuValaIleuValaIleuValaIleu 400
Db 1141 TCTGCCCATCTGGCCACCAAGCCCTCTCTCCCACTGGGACCCCTGGGAGCAAG 1200
Oy 401 TrpHisIleuValaIleuValaIleuValaIleuValaIleuValaIleuValaIleu 420
Db 1201 AAACACAGCATGAGCCAGAGAGCAAGCCAGCTGCGGGGCTGCAAGGCCCCCTG 1260
Oy 421 AlaSerGlyIleuValaIleuValaIleuValaIleuValaIleuValaIleuVal 439
Db 1261 GCTGCGAGAGAGCTCCACCTCATGGGCGGAGAGAGTGAAGAGTCTGCGGTGGTG 1320
Oy 440 ValTrpGluSerGlnValaIleuValaIleuValaIleuValaIleuValaIleu 459
Db 1321 GTGTACAGAGCAGGCTGGGAGGCGCTCAAGAGACAGAAACCATCAACCAACCA 1380
Oy 460 HisIleuHisIleuValaIleuValaIleuValaIleuValaIleuValaIleu 469
Db 1381 CATACCATCTATGAAAGTGGTTGACTCTCTGTGAACATTTCCCACTCAAGAGCA 1440

```

```

CURRENT APPLICATION NUMBER: US/09/993,966
CURRENT FILING DATE: 2001-11-27
PRIOR APPLICATION NUMBER: 60/252,884
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/291,109
PRIOR FILING DATE: 2001-05-16
PRIOR APPLICATION NUMBER: 60/325,571
PRIOR FILING DATE: 2001-10-01
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 1401
TYPE: DNA
ORGANISM: Mus sp.
US-09-993-966-6

Alignment Scores:
Pred. No.: 3,91e-196 Length: 1401
Score: 2171.50 Matches: 406
Percent Similarity: 91.01% Conservative: 19
Best Local Similarity: 86.94% Mismatches: 41
Query Match: 86.89% Indels: 1
DB: Gaps: 1

US-09-993-966-7 (1-470) x US-09-993-966-6 (1-1401)
Oy 1 MetGlyValLeuHisSerIleuProAlaIleuValCysIleuValaIleuValaIleuVal 20
Db 1 ATGGGAGAACTTCACTGAAAGCCGCGCGCTGTGCAAGCGCAGAGAGAGAGAGAGT 60
Oy 21 AsnSerPheAlaIleuValaIleuValaIleuValaIleuValaIleuValaIleuVal 40
Db 61 GACAGCTTCTCTGAAAGCGTGTGGTGGCAAGAAAGCAATCAAGAGTGAATGGAGAGG 120
Oy 41 GlnArgCysProGlyValaIleuValaIleuValaIleuValaIleuValaIleuVal 60
Db 121 CAGCGCTGTCAAGAGCAGCGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
Oy 61 ArgSerThrArgGluLeuValaIleuValaIleuValaIleuValaIleuValaIleuVal 80
Db 181 CGAGGCACTGGGAACTCGTGGGTGACACTTGAAGAGAGAGCTCGGTGAGAGAGAGAG 240
Oy 81 AsnAspPheArgLeuGluValaIleuValaIleuValaIleuValaIleuValaIleuVal 100
Db 241 GACGACTTCCCTTGAAGAGGCGCTGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Oy 101 AsnGluValLeuMetGluValaIleuValaIleuValaIleuValaIleuValaIleuVal 120
Db 301 GATGAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Oy 121 PheGluGluLeuGlnCysAspValaIleuValaIleuValaIleuValaIleuValaIleuVal 140
Db 361 TTGGAAGAGTCAAGTGTGATGCTCTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Oy 141 ThrLeuTrpAspPheAsnValaIleuValaIleuValaIleuValaIleuValaIleuVal 160
Db 421 ACTCTATATACCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Oy 161 HistHrIleTyrGluValValaAspSerSerValaAsnHisSerProThrSerSerIleuMet 180
Db 481 CATACCATCTATGAAAGTGGTTGACTCTCTGTGAACATTTCCCACTCAAGAGCA 540
Oy 181 LeuArgValIleuLeuThrValaIaProAspIleuSerIleuValaIleuVal 200
Db 541 CTCGGGGTGAAGCTCACCGTGGCTCTGACGGGAGCCAGATGAAGAGAGTCTTTTC 600
Oy 201 AsnGlnIleuAspLeuGlnSerIleuValaIaArgProArgIleuValaIleuVal 220
Db 601 AACCATACCATCTGACAGACCAAGAGCCCGAGCAGACCAACCCCGTGAAGAGCTG 660
Oy 221 ArgSerTrpGluValValaIleuValaIleuValaIleuValaIleuValaIleuVal 240
Db 661 CTTGGCTGGAGAAAGAGCAGAGCCCACTCAGTTTGAACCGGAGATCCCTTCGGTGGAG 720

```

QY 241 GlnSerGlyCysTyrHisHisCysValAspGluAsnIleGluArgArgAsnHisTyrLeu 260
 DB 721 CACCCAGACCTGCTACCACTGCGTGGATGAGAACTTGAAGGAAACCACTACTCTTA 780
 QY 261 AspLeuAlaGlyIleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValAla 280
 DB 781 GACCTGGCGGGGATAGAACTACACTCTCAGTTTGACCGGGAATCCCTTCGGTGGCC 840
 QY 281 GlnHisSerGluLeuProProArgThrSerAsnProThrArgSerArgSerHisIleGluPro 300
 DB 841 CAGAGCTCAGAGCTGCCCCCTCGAATCTCAACCCCACTGCTCTGCTCCAGAGCA 900
 QY 301 GlnAlaIleHisIleProHisArgGlyProGlnGlyValAspProAlaSerPheHisPhe 320
 DB 901 GAAGCTGCCACATCCCAACCGGAGGCCCAAGGTGTGACCCAGGCTCTTCCACTTC 960
 QY 321 LeuAspThrProIleAlaValSerGluLeuGlnGlnArgLeuArgGlyThrGlnAsp 340
 DB 961 CTTCACACCCCATTTGCGCAAGGCTACAGACTCCAGCAAGGCTCCGGGGCACTCAGGAT 1020
 QY 341 GlySerIleHisPheValArgSerProGlyAlaGlnGlyIleValSerValGlyValGlyHis 360
 DB 1021 GGGAGCAAGCACTTGTGAGGTCCCCCAAGGCCAAGGCAAGAAACATGGGTATGGGCGAC 1080
 QY 361 ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
 DB 1081 GGGGGCCAGAGGTGCAGAAACCAACCTTCCACTGGTATCCCAACCCCACTACTGTTCCCCC 1140
 QY 381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
 DB 1141 TCTGCCATCTGCGCAACAGCCCAAGCCCTTCTCCCACTCGGCAACCCCTGGGGCAACAG 1200
 QY 401 LysHisLysHisArgAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
 DB 1201 AAACACAGAGCATCAGCCCAAGAGAGCCAGGCGGAGCTGCGGGGCGCTGCAAGGCCCTTG 1260
 QY 421 AlaSerGlyGly---ProValLeuGlyArgGlnHisLeuArgGlyLeuProAlaLeuVal 439
 DB 1261 GCTGCGAGAGGCTCCACCGTATGGCGGGAGCGAGGTGAGGGAGCTGCTGCGGTGGTG 1320
 QY 440 ValTyrGluSerGlnAlaGlyGlnProValGlnArgHisGlnHisHisHisGlu 459
 DB 1321 GTGACAGAGCCAGGCTGGGCAAGCCGCTCCAGAGACACGAAACCATCACCAACCA 1380
 QY 460 HisHisHisHisTyrHisHis 466
 DB 1381 CATCACCACTATTCACAC 1401

RESULT 6
 US-09-730-989-1
 ; Sequence 1, Application US/09730989
 ; Patent No. US20020061552A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Williams, Lewis T.
 ; APPLICANT: Yan, Dong
 ; TITLE OF INVENTION: MAMMALIAN DISHEVELLED-ASSOCIATED PROTEINS
 ; FILE REFERENCE: PP-01657.002 / 200130.518
 ; CURRENT APPLICATION NUMBER: US/09/730.989
 ; NUMBER OF SEQ ID NOS: 29
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1
 ; LENGTH: 1401
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 US-09-730-989-1

Alignment Scores:
 Pred. No.: 9.38e-196 Length: 1401
 Score: 2167.50 Matches: 406
 Percent Similarity: 90.79% Conservative: 18
 Best Local Similarity: 86.94% Mismatches: 42

Query Match: 86.73% Indels: 1
 DB: 9 Gaps: 1
 US-09-993-966-7 (1-470) x US-09-730-989-1 (1-1401)
 QY 1 MetGlyLysLeuHisSerLysProAlaAlaValCysLysArgArgGluSerProGluGly 20
 DB 1 ATGGGGAATCTCACTCGAAGCCGGCCGCTGTGCAGAGCGAGGAGAACCCGGAAAGGT 60
 QY 21 AspSerPheAlaValSerAlaAlaATPAlaArgLysGlyIleGluGluTTPIleGlyArg 40
 DB 61 GACAGCTTGGCTGTAAAGCGTCTGTGGCGAAGAAAGGCATTCAGAGGTGAATCGGAGG 120
 QY 41 GlnArgCysProGlyGlyValSerGlyProArgGluLeuArgLeuAlaGlyThrIleGly 60
 DB 121 CAGCGCTGTCAAGCAAGCTCTTAAGACCCCTCACTGAGATTTGGCAGGCACTGTGGT 180
 QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80
 DB 181 CGAGGCACTGGGAACCTCGTGGGTGACATTTAGAGAGGCTCTGGGTGAGAGAGAG 240
 QY 81 AspAspPheArgLeuGluValAlaLeuProProGluLysThrAspGlyLeuGlySerGly 100
 DB 241 GACGACTTCCCCCTAGAGAGTGGCCCTGCGCTGAGAAAGATCGACAGCTTAAAGTATGGA 300
 QY 101 AspGluLysLysMetGluArgValSerGluProCysProGlySerLysGlnLeuLys 120
 DB 301 GATGAGAAAGATGAGAGACTGAGAGCACTGGCCAGGCTTCCAGAAAGCACTCAG 360
 QY 121 PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGlnGluTTPPhe 140
 DB 361 TTTGAAGACTACAGTGTGATGTCTGTGAGAGAGACAGCCGGCAAGTGGACTTTC 420
 QY 141 ThrLeuTyrAspPheAspAsnGlnLysValThrArgGluAspIleThrSerLeuLeu 160
 DB 421 ACTCTATATACATTCGACCAACATGGCAAGTACCGTGAAGACATTAACACTTGTCTG 480
 QY 161 HisThrIleTyrGluValAlaAspSerSerValAsnHisSerProThrSerSerLysMet 180
 DB 481 CATACCATCTATGAAAGGTGAGCTCTCTGTGAACATTCCTCCACATCAAGCAACACA 540
 QY 181 LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysArgSerValLeuVal 200
 DB 541 CTGGGGGTGAAGCTCACCGTGGCTCTGACGGAGCCAGAGTAAAGAGACGCTCTTTC 600
 QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeu 220
 DB 601 AACCATACCATCTGCAGACACAGAGCCCGGAGCAACACCAAAACCGCTGAGAGACTG 660
 QY 221 ArgSerTyrGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
 DB 661 CTGGCTGGAGAAAGAGCAAGCCAGCCCACTAGGTTCCAGGGTACAGCCCACTGGAG 720
 QY 241 GlnSerGlyCysTyrHisHisCysValAspGluAsnIleGluArgArgAsnHisTyrLeu 260
 DB 721 CAGCCAGACTGCTACCACTGCGTGGATGAGAACTTGAAGAGAAACCACTACTCTTA 780
 QY 261 AspLeuAlaGlyIleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValAla 280
 DB 781 GACCTGGCGGGGATAGAACTACACTCTCAGTTTGACCGGGAATCCCTTCGGTGGCC 840
 QY 281 GlnHisSerGluLeuProProArgThrSerAsnProThrArgSerArgSerHisIleGluPro 300
 DB 841 CAGAGCTCAGAGCTGCCCCCTCGAATCTCAACCCCACTGCTCTGCTCCAGAGCA 900
 QY 301 GlnAlaIleHisIleProHisArgGlyProGlnGlyValAspProAlaSerPheHisPhe 320
 DB 901 GAAGCTGCCACATCCCAACCGGAGGCCCAAGGTGTGACCCAGGCTCTTCCACTTC 960
 QY 321 LeuAspThrProIleAlaValSerGluLeuGlnGlnArgLeuArgGlyThrGlnAsp 340
 DB 961 CTTCACACCCCATTTGCGCAAGGCTACAGACTCCAGCAAGGCTCCGGGGCACTCAGGAT 1020

QY 341 G1SerIyShIshpHeValaArgSerProIyAlaGlnIyIySseValaG1yValaG1yHis 360
Db 1021 GGGAGCAGACATTGTGAAGTCTCCCAAGGCCCAAGGACAAAGACATGGATATGGGCAC 1080
QY 361 ValaIaArgG1yAlaArgAenIySProIeUeUgIyProAlaIleProAlaValaSerPro 380
Db 1081 GGGGCCAGAGCTGCAGAGACAGACCTCCACTGTATCCACCACTACCTCTCTCCCCC 1140
QY 381 SerAlaIleUeUaIaIaSerProAlaIeUeUProSerIeUaIaProIeUgIyHis 400
Db 1141 TCTGCCATCTGGCCACACAGCCCTTCTCCCACTCCGAGACCCCTGGGACAG 1200
QY 401 LysH1eUyShIshpHeValaArgIySseValaGlnIyIySseValaG1yValaG1yHis 420
Db 1201 AAACACAGCATGAGCCAGAGAGAGCCAGGAGCTCCGGGAGCTGCAGAGCCCTG 1260
QY 421 AlaSerG1yG1y---ProValaIeUgIyArgG1yHisIleUeUArgIySseValaIeUaI 439
Db 1261 GCTGAGAGAGCTCCACCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320
QY 440 ValIyG1ySerG1yAlaIyG1yGlnIyProValaGlnArgH1eUgIyHisIleUeU 459
Db 1321 GTGTACAGAGCAGAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380
QY 460 HisH1eUyShIshpHeValaArgSerProIyAlaGlnIyIySseValaG1yValaG1yHis 466
Db 1381 CATCACCACTATATACAC 1401
RESULT 7
US-10-087-192-176
Sequence 176, Application US/10087192
Publication No. US20020182586A1
GENERAL INFORMATION:
APPLICANT: Morris, David W.
APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
FILE REFERENCE: 52945200012
CURRENT APPLICATION NUMBER: US/10/087,192
PRIOR FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
NUMBER OF SEQ ID NOS: 2059
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 176
LENGTH: 2807
TYPE: DNA
ORGANISM: Mus musculus
US-10-087-192-176
Alignment Scores:
Pred. No.: 1,37e-187 Length: 2807
Score: 2085.50 Matches: 390
Percent Similarity: 90.67% Conservative: 18
Best Local Similarity: 86.67% Mismatches: 41
Query Match: 83.45% Indels: 1
Gaps: 13
US-09-993-966-7 (1-470) x US-10-087-192-176 (1-2807)

QY 81 AspAspPheArgIeUgIyValaIaIeUProProG1yUlySthraPheG1yIeUgIySerG1y 100
Db 183 GAGCACTTCCCTAGAAAGTGGCCCTCCCTGAGAAAGATCCGACAGCTAGATAGGGA 242
QY 101 AspG1yUlySthraPheArgIeUgIyValaIaIeUProProG1yUlySthraPheG1yIeUgIySerG1y 120
Db 243 GATGAGAGAGATGAG 302
QY 121 PheG1yUlySthraPheArgIeUgIyValaIaIeUProProG1yUlySthraPheG1yIeUgIySerG1y 140
Db 303 TTGGAAGCTACAGTGTGATGCTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 362
QY 141 ThrIeUyThraPheArgIeUgIyValaIaIeUProProG1yUlySthraPheG1yIeUgIySerG1y 160
Db 363 ACTTATATACCTTGCAGCAATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 422
QY 161 HisH1eUyShIshpHeValaArgSerProIyAlaGlnIyIySseValaG1yValaG1yHis 180
Db 423 CATACATCTATGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 482
QY 181 LeuArgValaIySseValaIaIaProAspG1ySerIleUeUArgIySseValaIeUaI 200
Db 483 CTGCGGAGAGCTCACCGTGGCTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 542
QY 201 ArgG1yAlaIeUgIySseValaIaIaProAspG1ySerIleUeUArgIySseValaIeUaI 220
Db 543 AACCATACCATCTGCAGACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 602
QY 221 ArgSerTrpG1yUlySthraPheArgIeUgIyValaIaIeUProProG1yUlySthraPheG1yIeUgIySerG1y 240
Db 603 GGTGGTGGAG 662
QY 241 GlnSerG1yUlySthraPheArgIeUgIyValaIaIeUProProG1yUlySthraPheG1yIeUgIySerG1y 260
Db 663 CAGCAGAGCTGATACACCATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 722
QY 261 AspLeuAlaIyIleUeUArgIySseValaIaIaProAspG1ySerIleUeUArgIySseValaIeUaI 280
Db 723 GACCTGGCGGAGATGAGAACTACGCTCAGATTTGAGAGGAGATCCCTTCCGTGGCC 782
QY 281 GlnIySseValaIaIaProAspG1ySerIleUeUArgIySseValaIeUaI 300
Db 783 CAGAGTCAAGAGCTGCTCCCAATCTCCAGCCCACTGCTCTCCCAAGAGCA 842
QY 301 GlnAlaIleIshIshpHeValaArgSerProIyAlaGlnIyIySseValaG1yValaG1yHis 320
Db 843 GAGGCTCCCAATCCACACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 902
QY 321 LeuAspThrProIleAlaIySseValaIaIeUProProG1yUlySthraPheG1yIeUgIySerG1y 340
Db 903 CTGACACCCCATTTGCAAGGATCAGAGCTCAGAGAGGCTCCGAGGAGGAGGAGGAGG 962
QY 341 G1ySerIyShIshpHeValaArgSerProIyAlaGlnIyIySseValaG1yValaG1yHis 360
Db 963 GGGAGCAGACATTGTGAAGTCTCCCAAGGCCCAAGGACAAAGACATGGATATGGGCAC 1022
QY 361 ValaIaArgG1yAlaArgAenIySProIeUeUgIyProAlaIleProAlaValaSerPro 380
Db 1023 GGGGCCAGAGCTGCAGAGACAGACCTCCACTGTATCCACCACTACCTCTCTCCCCC 1082
QY 381 SerAlaIleUeUaIaIaSerProAlaIeUeUProSerIeUaIaProIeUgIyHis 400
Db 1083 TCTGCCATCTGGCCACACAGCCCTTCTCCCACTCCGAGACCCCTGGGACAG 1142
QY 401 LysH1eUyShIshpHeValaArgIySseValaGlnIyIySseValaG1yValaG1yHis 420
Db 1143 AAACACAGCATGAGCCAGAGAGAGCCAGGAGCTCCGGGAGCTGCAGAGCCCTG 1202
QY 421 AlaSerG1yG1y---ProValaIeUgIyArgG1yHisIleUeUArgIySseValaIeUaI 439
Db 1203 GCTGAGAGAGCTCCACCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1262

QY 231 ----- 231
 Db 94219 GAGGAGGGCTTAGGAGCCCCAGCCCTGCAGGGGAAAAAAGAGCCTTGAGCATC 94278
 QY 231 ----- 231
 Db 94279 TCTCCCAACCCCTCATGACACCTGTGTAAAGTCAACATGATGTGACACACACTCA 94338
 QY 231 ----- 231
 Db 94339 CACACTGTGAGGCCCATCTGGGCGTAGGGCAGATCATGGGTCTGCTGGCCAAAC 94398
 QY 231 ----- 231
 Db 94399 CCAACAGTTTAAAGGGATTCTGCTGCCACACCCCTGAGGAAACAGGCGAGGGTTAGG 94458
 QY 231 ----- 231
 Db 94459 CCGGCTTACCTGACGTGGGCAATATGGCTTGAAGCCTGGGCACTTCTAGAAAT 94518
 QY 231 ----- 231
 Db 94519 TGAACATACTCTCAGCTTTGCTTTGAACCTGTCAGAGCAAAACCCGTCACACTCT 94578
 QY 231 ----- 231
 Db 94579 CCCAAGAGCTGCGGGAGGTGCGGCTGTGGCTGTGTCTATGTCTGTCTCCATCGG 94638
 QY 231 ----- 231
 Db 94639 CAGGAGGGAATGAGGAGCAGTTCGTATAGAGACTGTTCTCCCACTCTCCAAA 94698
 QY 231 ----- 231
 Db 94699 GGGCAAGAAATGATGAGAGGTCACAGATTCTTGTGGGGCTTCTAGAGCTA 94758
 QY 232 ----- Arg 232
 Db 94759 TGGCTTTGCCCCCACTGTGTGTGTATCCCACTTGCCTCCCTGCGGTAGG 94818
 QY 233 PheGlnGlyAspSerArgLeuGlnSerGlyCysTyrHisIleCysValAspGluAsn 252
 Db 94819 TTCAGAGGTACAGCCCCCTGAGGCACTGTGCTGCTACCACTTCCTGATGAGAAC 94878
 QY 253 IleGluArgGlyArgHisIleTyrLeuAspLeuIleGlyTyrHisSerGlnPhe 272
 Db 94879 ATCGAGGAGGAAACCACTACTTAGATCTCGCGGATGAAAACTACAGTCCCAATTT 94938
 QY 273 GlyPro----- 274
 Db 94939 GGGCTGTGTAGGGACTCAAGCA CCGTCAATGGCGATGAGGGCAGGGCGTGGAC 94998
 QY 274 ----- 274
 Db 94999 GGGCCAGCGCGGCGCTGCGGGTGTGTTCCTGCTACCCCAAGCTTCGGTTAAGCAACTA 95058
 QY 274 ----- 274
 Db 95059 TTATGGGACAGTCAAAAGACTTGTGAGAAAGTAGAGTTAAATGTTCAAAATTA 95118
 QY 274 ----- 274
 Db 95119 AGTAATACATGACGAATTCGAA CAAATTCMAATAGTATCTAAAGGTTAGCACCCAATG 95178
 QY 274 ----- 274
 Db 95179 ACAAAACAAGTTCTGTGCCCCATCCAGGCCAACAGCCGTGTTCCGGGGCAACAC 95238
 QY 274 ----- 274
 Db 95239 TTGACCTCTCTGCTGTTTCTTCAGCAAGAACTCGGTTTCCCTTAGTACAGACTA 95298
 QY 274 ----- 274

Db 95299 CTGTGCTTTATACGCTTGAATATGTGAACCTTTGCTCAATTTCACTCCCAACTC 95358
 QY 274 ----- 274
 Db 95359 CTTTCTCCCTTTCTCTTCATATGTTTAATTCCTATGATTAAGTACTTGTATGTTTTTA 95418
 QY 274 ----- 274
 Db 95419 GCATATACCTTAGAGTTCTTTCTTGCCCAATGACCTTAGATAGTTTTCGGGATCTGTC 95478
 QY 274 ----- 274
 Db 95479 TGCCATGTCTTTCTCGGCTCCCTCTCGGAGATATTTTGAAGACTTACCAAGAAC 95538
 QY 274 ----- 274
 Db 95539 AGTTCTTCCACTTAGAGATGGGAAAGTTCCATTTGAGAGTTTTCCTGAAGGAGT 95598
 QY 274 ----- 274
 Db 95599 GAGGGCAGTCAAGGCAATTGGGGGTAGCTGTGTTTGAGAACTGGGGCGGGGTGA 95658
 QY 274 ----- 274
 Db 95659 TGTCTGGGTATAGCCGCAAGCCCAAGACACAGTAGTGCTCATTTAAATGCTGTGAAT 95718
 QY 275 ----- GlySerProSerValAlaGlnIleSerGlyLeuProPheArgHisSerAsn 291
 Db 95719 TGGTTCCTAGGCTCCCTTCCGTGTGGCCAGAACTGAGATGCCCCCGCACTCCAT 95778
 QY 292 ProThrArgSerArgSerHisGluProGluAlaIleHisIleProHisArgIleProGln 311
 Db 95779 CCCACTGATCTGCTCCCATGAGCCGGAAGCATCCACATCCACACCAAAAGCCCCAA 95838
 QY 312 GlyValAspProIleSerPheHisIleLeuAspThrProIleAlaIleValSerGluLeu 331
 Db 95839 GGGGTGACCCGGGCTCTCTTCCATCTTCACTTCAAGAACCCCAATCCGCAAGTCTCAGACTC 95898
 QY 332 GlnGlnArgLeuArgGlyThrGlnAspGlySerIleHisIlePheValArgSerProIleVal 351
 Db 95899 CAGCAACGCTCCGGGGCAACCCAGAACGGAGAGACACTTGTGAGTCCCCCAAGGCC 95958
 QY 352 GlnGlyValSerValGlyValGlyHisValAlaArgGlyValArgGlnIleProIleu 371
 Db 95959 CAGGGCAAGAGTGTGGGTGTGGGCCACGTGGCCAGAGGGCAAGAAACAAGCCCCCTG 96018
 QY 372 GlyProAlaIleProIleAlaValSerProSerAlaHisIleValAlaSerProAlaLeu 391
 Db 96019 GAACCCGCAATCCCTGCGGTGTCCCTCCGCGCACCTGCTCCAGCCCGGCTTCCTC 96078
 QY 392 ProSerLeuAlaProLeuGlyHisIleValHisIleValHisIleValAlaValSerGln 411
 Db 96079 CCTCTCTAGCCCCCTCCGGGCAACAAGGACCAAGCAACGAGCCCAAGAGACGACAG 96138
 QY 412 GlyCysArgGlyLeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlyArgHis 431
 Db 96139 GGCCTGCGGGGCTGTGAGGACCACTGGCTCAGGTGCTCTGCTGCGGGCGGAGAC 96198
 QY 432 LeuArgGlyLeuProIleAlaValIleTyrGlnSerGlnAlaGlyIleProValGlnArg 451
 Db 96199 CTGCGGAGCTGCGCCCTTGTGTGTATGAGAGCGAGCGGCGAGCGGCTCAGAGA 96258
 QY 452 HisGluHisIleHisIleHisIleGluHisIleHisIleTyrHisIleHisIlePheTyrGlnThr 470
 Db 96259 CATGAGCACACACACACATGAATCAACCACTTACCAACACTTACAGACAC 96315

RESULT 9
 US-10-087-192-175
 ; Sequence 175, Application US/10087192
 ; Publication No. US20020182586A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Morris, David W.

APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
FILE REFERENCE: 52945200122
CURRENT APPLICATION NUMBER: US/10/087,192
PRIOR FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 2059
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 175
LENGTH: 98474
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(98474)
OTHER INFORMATION: n = A,T,C or G
US-10-087-192-175

Alignment Scores:
Pred. No.: 7.57e-90 Length: 98474
Score: 1076.00 Matches: 278
Percent Similarity: 29.11% Conservative: 18
Best Local Similarity: 27.34% Mismatches: 39
Query Match: 43.06% Indels: 683
DB: 13 Gaps: 6

US-09-993-966-7 (1-470) x US-10-087-192-175 (1-98474)

QY 129 SerMetGluluaPserAArgIngluTrpThreThreLeuTyraePheAspAsn 148
DB 83990 TCTCAGGTTGCTGATGGCAGGCGATCGTG-----CTAAAGCTTTCG 84031
QY 149 GLYeVal-----ThArgluAspIleThSerleuLeuH1sThIleTygluVal 166
DB 84032 GGGGTCTGCTTCTCTCCCAAGAGACATTCACAGCTTGCTGCATACCTATGAAATG 84091
QY 167 ValAspSerValaenH1sSerProThSerSerlyMetLeuValIyLeuThr 186
DB 84092 GTTGACTCTCTGGAACCATTCCTCCCAATCAAGAACACACACGCGGTGAAGCTCAC 84151
QY 187 ValaIaProAspGlySerGlnSerLyArgSerValIeUValaen----- 201
DB 84152 GTGGCTCTTGACGGGAGCCAGAGTAAGAGAGCGTCTTTCAACATACCGGTAAGAG 84211
QY 201 ----- 201
DB 84212 GCGGAGACAGTCCCACTATGGGAAAGACAGGAGATCTAAGTCAGCTGGAGGTTG 84271
QY 201 ----- 201
DB 84272 GAGAGACCAGACTACATAGACAGTTGCTTGTGTGTACTGCAAGGCTTCAGCAA 84331
QY 201 ----- 201
DB 84332 ACCCTTAAGAGCTAAGCCTCTGGGAGGAGAGAGGAGCGGAATCAAGATGGAAGG 84391
QY 201 ----- 201
DB 84392 CCCAGTGAAGCAAGGATAGAGTTGAGACTGAGAGAGATGCAAGGCCAGGCTTAGAGG 84451
QY 201 ----- 201
DB 84452 GCAGCTTATGAATCTGGAGAGGTGTGGGACAGGCAACTAGTTTGCTCTTGCTGGGA 84511
QY 201 ----- 201
DB 84512 ACCCTGTGAGAGCTGTGTGAATCTATTGAGTGGCTTTTGAGCACACAAATGCCAGGA 84571
QY 202 -----GlnaIaasp 204

DB 84572 AGCCTGGCCGAGTTGGAGAGGCACTGACAGTCTCTTTGACTTCTGGACTCAGAT 84631
QY 205 LeuGlnSerAlaArgProAlaIaGluThrLyArgProThrGluAspLeuArgSerTyglu 224
DB 84632 CTCGAGAGCACAAAGGCCCCCAGAGACACCAACCCGCTGAGAGCTGTGGCTGGAG 84691
QY 225 LyLeuGlnArgIaProLeu----- 231
DB 84692 AAGAGCAGCGAGCCCACT-CAGGTAAGACATGCTGTGCCCTGAAACATGATTGAG 84750
QY 231 ----- 231
DB 84751 CGCAGCTGATATGTACATCTGAAGACTTTGAGTAGCAGTTCCATCTGGGTCTCT 84810
QY 231 ----- 231
DB 84811 GCAGGGCCATGATTTATTTTCAGAAAGACAGATTTACAAATCAGACTCTAAATTACA 84870
QY 231 ----- 231
DB 84871 AACAAAACAGAAAAAAATCAAAACCTAGATATAGCTCATCAGTGTATTCATATA 84930
QY 231 ----- 231
DB 84931 AAATTTTAGATTTTCTCATGTGTGTTGATTCAGCATGATATGGGTAAATTCGTGATG 84990
QY 231 ----- 231
DB 84991 TATTTCCATGTAGATATAGTAGGTAGACACTTGGGGTCTAAACACTGGGATTCTGC 85050
QY 231 ----- 231
DB 85051 TGGGAGATATGTGATTAAGCATACACAGACAGAGGACMAAACTTCCAGGCTCTGCCC 85110
QY 231 ----- 231
DB 85111 TTAAGATGGCTGCTGTGTGAGTGGCTATGGGMAAACCATTGTGAGGTGCTCAG 85170
QY 231 ----- 231
DB 85171 AATGTGATTAAGACTTATATAGTCTGACAGTTATCTGTCCGTTGTCACTTCTGCATACA 85230
QY 231 ----- 231
DB 85231 CAGAGGAAAGAGGATGTTTAAAGAAAGCACTTAACAGACAGAGGAGGAGCAAGCA 85290
QY 231 ----- 231
DB 85291 AGCGTAGATTTTCAACATCTGTGCCAAAGAGCGTGGCCAGAGATCGGCTAGTCTT 85350
QY 231 ----- 231
DB 85351 CATGTAGCGTGTATTAAGACAGACCATGTAAAGTGGCAGTGGGAGGCTGGCTGAGAT 85410
QY 231 ----- 231
DB 85411 GGGTGGGCTGGGGGCAATTGAACAGACTTTCAGTGGTGAAGGAGATGTGTGGGCACT 85470
QY 231 ----- 231
DB 85471 GGGGAATGAAGGACAGACAGGATTCCTGCTCAGAGCCCTTCTCCAAATGCACAGAA 85530
QY 231 ----- 231
DB 85531 ACGTGAAGATGATGAAGCTGCCGGTTTGCCCTCAGCTTCCACAGAGCTCAGGGT 85590
QY 232 -----ArgPheGlnIyaPserArgLeuGlu 240
DB 85591 CTGTGCTTCTTACATATATGCTCTTTTGTACACAGGTTCCAGGTTGACAGCACTGTGAG 85650
QY 241 GlnSerGlyCystrYrH1sH1sCyValaAspGluAsnIleGluArgArgAsnH1sTyLeu 260
|||

```

Db      85651 CAGCAGACTGCTACCAACCATGCGTGGATGAGAAACATTGAGAGAGAAACACTACTACTA 85710
Qy      261 AspleuAla1y1leg1uan1y1ThSer1nphel1y----- 273
Db      85711 GACCTGGCGGGATAGAGAACTACACGTCTCAGTTTGACCGGATAGGGGCTGTGT 85770
Qy      273 ----- 273
Db      85771 GGGCCATAGTCTGACACACAGTGTGGTATGTAACCATGATCCTGAGGACTGGGAG 85830
Qy      273 ----- 273
Db      85831 ATAGTAGCTTTTCTACTGACAGGTGAGGCTGTCTCAGAGCCCTTAACAAAGACTG 85890
Qy      273 ----- 273
Db      85891 AAGACATTTTCAAAATCAAAATGTAATACTTAGATTAAGCTGGCAAAATCCAGAGCAA 85950
Qy      273 ----- 273
Db      85951 GAGAGACTGAGAGTCCAGCAAAAGCAGTAGACCCGCTCAAGCCAGCCCGGGGCC 86010
Qy      273 ----- 273
Db      86011 CACATGACCCCAAGCCCAAGCCCAAGCCCAAGCCCAAGCCCAAGCCCAAGCCCAAGCCCA 86070
Qy      273 ----- 273
Db      86071 AGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCCCCAGCGGCTGTGTCAGAGAAATGT 86130
Qy      273 ----- 273
Db      86131 GCTGCTGTGCAAGCTGTGTGCCCTGACCTGTGTGCCAGTCTGTCTTCTTAATGTT 86190
Qy      273 ----- 273
Db      86191 GAGCTCTCTGAGATTGACCCGTGATTTCCCTTATCTATTAGAGGCTGTGCCACT 86250
Qy      273 ----- 273
Db      86251 CCTTAGTCCCTTCTCTCTCTCTCCTCAGCAGATGGAGACATGTTTCTCCCTA 86310
Qy      273 ----- 273
Db      86311 CAGGACTGTAGGGCATTAGAGGGCTGTAGCTCAGACAGAGTCAATGCTCCGTGTA 86370
Qy      274 -----Pro 274
Db      86371 ACCCAGGGTACCCAGTGCACACAGTGGTCTCATGAAGTATCTGCGATTGCTTCT 86430
Qy      275 -glySerProSerValAla1n1y1Ser1n1y1Leu1Pro1n1y1ThSer1n1y1Pro1n1y1 294
Db      86431 AGGATCCCTTCCGTGGCCCAAGAGTCAAGAGTGTCCCTCGATCTCAACCCCACTCG 86490
Qy      294 gSer1n1y1Ser1n1y1Leu1Pro1n1y1ThSer1n1y1Pro1n1y1Val1n1y1 314
Db      86491 CTCTCGTCTCCAGAGCAGAGTGTCCCAATCTCCACCGGAGCCCAAGGTGTGGA 86550
Qy      314 pPro1n1y1Ser1n1y1ThSer1n1y1Leu1Pro1n1y1ThSer1n1y1Pro1n1y1Val1n1y1 334
Db      86551 CCAAGGCTCTTCCACTCTTTCAGACCCCACTTTCAGAGGATCAAGCTCCAGCAAG 86610
Qy      334 gLeu1n1y1Th1n1y1n1y1Ser1n1y1ThSer1n1y1Pro1n1y1Val1n1y1 354
Db      86611 GCTCCGGGACATCAGATGGAGCAAGCACTTGTGAGGTCCCAAGGCCCAAGGCA 86670
Qy      354 sSer1n1y1Val1n1y1ThSer1n1y1ThSer1n1y1Pro1n1y1Val1n1y1 374
Db      86671 GAACATGGATAGGCGCAGAGGCGCAGAGGCGCAGAGGCGCAGAGGCGCAGAGGCGCAG 86730
Qy      374 a1e1Pro1n1y1Ser1n1y1ThSer1n1y1ThSer1n1y1Pro1n1y1Val1n1y1 394
Db      86731 CACCATCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 86790

```

```

Qy      394 uAlaProLeuG1yH1s1y1Ser1n1y1ThSer1n1y1Pro1n1y1Val1n1y1 414
Db      86791 GGCACCCCTGGGACAGAAACAGACATCGAGCCAGAGAGCCAGCGGAGCTGCCG 86850
Qy      414 gG1yLeuG1nAlaProLeuAlaSer1y1y1y1ProVal1n1y1Arg1n1y1LeuAr 433
Db      86851 GGGCTTCAGAGGCGCCCTGTGCTGAGAGGCTCCACCGTCAATGGGCGGAGCAGTGAG 86910
Qy      433 gG1yLeuProAlaLeuVal1y1y1y1Ser1n1y1Ala1y1G1nProVal1n1y1Arg1n1y1 453
Db      86911 GGAAGTCTCGCTGGCTGTGTGTGTAACAGAGCCAGCTGGGCGAGGCTCCAGAGACAGA 86970
Qy      453 uH1s1y1H1s1y1H1s1y1H1s1y1H1s1y1H1s1y1H1s1y1H1s1y1H1s1y1 469
Db      86971 ACACCATCACACACAGACAGACATCAACCATTAATACACTTCTATCAG 87019

RESULT 10
US-09-764-891-5902
; Sequence 5902, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; PRIOR APPLICATION DATA: 2001-01-17
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 5902
; LENGTH: 2379
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-5902

Alignment Scores:
Pred. No.: 4,49e-91 Length: 2379
Score: 1067.00 Matches: 257
Percent Similarity: 58.08% Conservative: 34
Best Local Similarity: 51.30% Mismatches: 100
Query Match: 42.70% Indels: 110
DB: 10 Gaps: 16

US-09-993-966-7 (1-470) x US-09-764-891-5902 (1-2379)
Qy      29 TrpAlaArg1y1Ser1n1y1ThSer1n1y1Pro1n1y1Val1n1y1 48
Db      82 TGGGCGATAGAGGCGAGGCGCTGTGACCGGCCAGCGGCGGCGGTGTGTCTCAC 141
Qy      49 G1yProArg1n1y1LeuAla1y1Th1y1y1Arg1n1y1ThArg1n1y1LeuAla1y1 68
Db      142 TGCATCCCGAGGCTTGGTAAAGCAACTATTATAGGACAGTCAAGA---CTTCTTGA 198
Qy      69 Asp-Val1y1LeuArg1n1y1ThSer1n1y1ThSer1n1y1Pro1n1y1Val1n1y1 88
Db      199 GAAAGTACGCTTAATAATGTTTCAAAATTAAG---TAATACATGACAGCAAAATCGA 252
Qy      88 ala1y1ProG1n1y1ThSer1n1y1ThSer1n1y1ThSer1n1y1Pro1n1y1Val1n1y1 108
Db      253 ACAAAATCAATATGATTAAGGCTTGAACCAATGACAAACAAACAGTTTGTGCC 312
Qy      108 lSerG1nProCy1s-----ProG1n----- 114
Db      313 ATCCAGCCCAACCAAGCCCTGTTCCTGGGGAACCACTTGAACCTCTCTGCTGTTC 372
Qy      115 -Ser1y1y1G1n1y1ThSer1n1y1ThSer1n1y1Pro1n1y1Val1n1y1 134
Db      373 TTCAGCAAAACCTCGTTTCCCTTAAGTACAGACACTACTGTGTCTTAA----- 421
Qy      134 rArg1n1y1Trp1n1y1ThSer1n1y1ThSer1n1y1Pro1n1y1Val1n1y1 154
Db      421 ----- 421

```

```

Qy 154 uAaplleThrSerleuLeuHsiThrIleTyrgluValValAspSerSerValaenHisse 174
Db 422 -----TCAGCTTGACATATGTTAATCTTCTGCTCAATTTTCAC 459
Qy 174 rProThrSer----- 177
Db 460 TCCCAACAACCTCTTCCCTTCTCTCATATGTTAAATCTCTATAGTACTTCT 519
Qy 178 -----SerlyMetLeuArgValLysLeuThValAlaProAsp----- 190
Db 520 GATATGTTTTAGCAATATACTTAGAGTT---CTTCTTGCCCCAGTAACTTAGATAGTT 576
Qy 191 -----GlySerGlnSerlyAspSerValleuVa 200
Db 577 TTCGGGGTCTCTGCTGCATCTGTCTTCTGAGCTCTCTCCGAG--ACTATTTT-- 631
Qy 200 laenGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIAspProThrgluAspLe 220
Db 632 -----GAAGACCTA-----CCAAAGAACAGTTCTTCCA-----CT 663
Qy 220 uArgSerTrpGluLys-----LysGlnArgAlaProLeuArg----- 232
Db 664 TAGCAATGGGAAAGTTCATTTTCAGAGAGTTTCTGCAAGGCAGTGAGGGGAGTCAG 723
Qy 233 ----PheGlnGlyAspSerArgLeuGlnSerGlyCysTrpHsiAsyValAspG1 251
Db 724 GGCAATTGGGGGAGTCTGCTGTTTGGAGAACCTGGGGCG-----GGGGTGAATGTC 774
Qy 251 uAsnIleGluArgHsiAsnHsiTyrlLeuAspLeuAlaGlyTrIleGluAsnTyrlThrSerG1 271
Db 775 TGGGGTATAGCGCAAGCCCAAGCAC-----ACAGTAGTGCTCATTAATATGTCGTGTA 828
Qy 271 nPheGlyPro-GlySerProSerValAlaGlnLysSerGluLeuProProArgHThrSerA 291
Db 829 ATTGTTCTCAGAGCTCCCTTCCTGCGCCAGAACTGCAAGACTCCCCCGGCACCTTCA 888
Qy 291 snProThrArgSerArgSerHsiGluProGluAlaIleHsiIleProHsiArgHyspProG 311
Db 889 ATCCCACTCATCTCGCTCCCATAGCCGAGAACCATTCACATCCACACCGAAAGCCCC 948
Qy 311 lngIValAspProAlaSerPheHsiPheLeuAspTrpProIleAlaLysValSerGlu 331
Db 949 AAGGCGTGAACCCGGCTCTCTTCCACTTCTTGACACCCCAATGCCAAGGTCTCAAGGC 1008
Qy 331 euGlnGlnArgLeuArgGlyThrGlnAspGlySerLysHsiPheValArgSerTrpLysA 351
Db 1009 TCCAGCAACGGCTCCGGGGCACCCAGAGCGGAGCAAGCACTTGTGAGTCCCCCAAGG 1068
Qy 351 laGlnGlyLysSerValGlyValGlyHsiValAlaArgGlyValaArgAsnLyspProIol 371
Db 1069 CCCAGGCGAAGAGTGGGTGGGTGGCCACGTGGCCAGAGGGGCAAGAAACAGCCCTTC 1128
Qy 371 euGlyProAlaIleProAlaValSerProSerAlaHsiLysLeuAlaLysProAlaLeuL 391
Db 1129 TGGGACCCGGCATCTCGGGGTGTCCCTCGCCACTTGCTGGCCAGCCCGGCTCC 1188
Qy 391 euProSerLeuAlaProLeuGlyHsiLysLysHsiLysHsiArgAlaLysGlnSerGlnG 411
Db 1189 TCCCTCTCTAGCCCTCGGGGCAAGAGCAAGCAAGCAAGCAAGCAAGAGGCCGCGC 1248
Qy 411 lngIYcYAspGlyLeuGlnAlaProLeuAlaSerGlyLysProValLeuGlyArgGluH 431
Db 1249 AGGGCTGCGGGGCTGCAAGCAACCTGCTCAGGTGGCTCTGCTGGGGGGAGGC 1308
Qy 431 iAsLeuArgLysLeuProAlaLeuValTyrgluSerGlnAlaGlyGlnProValGlnA 451
Db 1309 ACCTGCGGGAGCTCGGCTTGGTGGTGTATGAGAGCAAGCGGGGAGCGGATCCAGA 1368
Qy 451 rghAsGluHsiHsiHsiHsiHsiHsiHsiHsiHsiHsiHsiHsiHsiHsiHsiHsiHsi 470
Db 1369 GACATGAGCAACCAACCATGAACTACCAACCACTTACCAACCACTTACCAAGACA 1427

```

```

RESULT 11
US-09-764-891-5904
; Sequence 5904, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; PRIORITY FILING DATE: 2001-01-17
; Prior application data removed - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5904
; LENGTH: 2379
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-891-5904

Alignment Scores:
Pred. No.: 4,49e-91 Length: 2379
Score: 1067.00 Matches: 257
Percent Similarity: 58.08% Conservative: 34
Best Local Similarity: 51.30% Mismatches: 100
Query Match: 42.70% Indels: 110
DB: 10 Gaps: 16

US-09-993-966-7 (1-470) x US-09-764-891-5904 (1-2379)

Qy 29 TrpAlaArgLysGlyIleGluGluTrpIleGlyArgGlnArgCyspProGlyGlyValSer 48
Db 82 TGGGCGATGAGGAGGCGCTGCTGAGACGGGCAAGCGCGCGCTGCGGTGTTGCAC 141
Qy 49 G1yProArgGlnLeuArgLeuAlaGlyThrIleGlyArgSerThrArgGluLeuValGly 68
Db 142 TGTACCCCAAGGCTTCGTGTAAGCAACTATTATGGACAGGTCAAGA---CTTCTGGA 198
Qy 69 Asp-ValLeuArgAspThrIleuSerGluGluGluGluAspAspPheArgLeuGluValAl 88
Db 199 GAAGTACCTTAATAATGGTTTCAAAATTAAG-----TAATACATGACGAAATCGA 252
Qy 88 aleuProGluLysThrAspGlyLeuGlySerGlyAspGlyLysLysMetGluArgVa 108
Db 253 ACAAAATCAATAGTATCTAAAGCTTAGCACCAATGACAAACAAACAGTTTGTGCCCC 312
Qy 108 lSerGluProCys-----ProGly----- 114
Db 313 ATCCCAACCCCAACCCAGCCCTGTTCCTGGGGCAACCACTTGACCTTCTGCTGTTTC 372
Qy 115 -SerLyLysGlnLeuLysPheGluGluLeuGlnCysAspValSerMetGluuAspse 134
Db 373 TTCCAGAAAGAACTCGTTTCCCTAAGTACAGACTACTGTGCTTTA----- 421
Qy 134 rArgGlnGluTrpThrPheThrLeuTyraSpPheAsnAsnGlyLysValThrArgG1 154
Db 421 ----- 421
Qy 154 uAaplleThrSerleuLeuHsiThrIleTyrgluValValAspSerSerValaenHisse 174
Db 422 -----TCAGCTTGACATATGTTAATCTTCTGCTCAATTTTCAC 459
Qy 174 rProThrSer----- 177
Db 460 TCCCAACAACCTCTTCCCTTCTCTCATATGTTAAATCTCTATAGTACTTCT 519
Qy 178 -----SerlyMetLeuArgValLysLeuThValAlaProAsp----- 190
Db 520 GATATGTTTTAGCAATATACTTAGAGTT---CTTCTTGCCCCAGTAACTTAGATAGTT 576
Qy 191 -----GlySerGlnSerlyAspSerValleuVa 200
Db 577 TTCGGGGTCTCTGCTGCATCTGTCTTCTGAGCTCTCTCCGAG--ACTATTTT-- 631

```

OY	200	AaenGlnAlaAerLeuGlnSerAlaAtrProAtrgAlaGlnPthLyrProThnGluAer	220
Db	632	-----GAAAGCTA-----CCAAAGAACAGTTCTTCCSA-----CT	663
OY	220	uAtrSerTrpGlnLyr-----LyeGlnAtrgAlaProLeuAtrg-----	232
Db	664	TAGCAGATGGGAAGTTCATTTCAAGAGAGTTTCTCGAAGGCAAGTAGAGGGAGCTAG	723
OY	233	----PheGlnGlyAerSerAtrgLeuGlnInserGlyCyvTyGlnAhiCyAValAerG1	251
Db	724	GGCATGGGGGGGTACTCTGTGTTTTGAGAACTGGGGGGCG-----GGGGTATGTC	774
OY	251	uAenIleGlnTygAaGdaAenIhiTygLeuAerLeuAhiGlyIleGluAenTyThSerG1	271
Db	775	TGGGGTATAGCGCAAGCCCAAGCCAC-----ACAGTAGGTGTCATTAATGTCGTGTTA	828
OY	271	nPheGlyPro-GlySerProSerValAlaGlnLysSerGluLeuProProAtrThnSerA	291
Db	829	ATTGGTTCCTAGAGCTCCCTTCCTCGGGCCCAAGTCAAGATGAGTCCCCCGCACTTCA	888
OY	291	aProThrAtrgSerAtrgSerhiGlnProGluAlaIlehiAerProhiAtrgLyvProG	311
Db	889	ATCCCACTCGATCTGCTCCCAAGCGGAAGCCATCCATCCACACCGAAAGCCCC	948
OY	311	InGlyValAerProAlaSerPhehiAerPheLeuAerThrProIleAlyValSerGln	331
Db	949	AAGGCTGAGACCGGCGCTCTTCACATTCCTTGACACCCCAATCGCAAGGTCTCAAGAC	1000
OY	331	euGlnGlyAtrgLeuAtrgGlyThrGlnAerG1SerLyvhiAerPheValAtrgSerProLyvA	351
Db	1009	TCCAGAACCGGCTCCGGGGGACCCCAAGACGGAGACGACATTTGTAGGTCCCGAAGG	1066
OY	351	IaGlnGlyLysSerValGlyValGlnhiAValAlaTrgGlyAlaAtrgAenLyvProProL	371
Db	1069	CCAGAGGCAAGTGTGGGTGTGGGCAAGTGGCCAGAGGGGCAAGAACAGCCCTCTC	1122
OY	371	eugLyProAlaIleProAlaValAerProSerAlahiLeuAlaAerProAlaLeuL	391
Db	1129	TGGGACCGGCATCCCTCGGTGTCTCCCTCCGCCACCTGCTGCCACCGCGGCTCTC	1188
OY	391	euProSerLeuAlaProLeuGlnhiAlyvLyvhiAerGlyAlaLyvGlnSerGln	411
Db	1189	TCCCTCCCTTAGCCCCCTCGGGACAAAGAACACAGCACCGAGCCAAAGAGAGCGAC	1244
OY	411	InGlyCyvAtrgGlyLeuGlnAlaProLeuAlaSerGlyG1ProValLeuGlyAtrgGln	431
Db	1249	AGGGGTCCGGGGGCTCGAGGACACAGCTAGGCTCAAGTGGCCCTCTCGGGGGCGAGAC	1300
OY	431	IeLeuAtrgGlnLeuProAlaLeuValTyTrgLysSerGlnAaGlyGlnProValGlnA	451
Db	1309	ACCTGCGGAGACTGCCCTTGGTGTGTGTATGAAGCCAGAGCCGGGAGCGGTCACGA	1366
OY	451	rGhiAgluAhiAhiAhiAhiAhiAgluAhiAhiAhiAhiAhiAhiAhiAhiAhiAhiAhiA	470
Db	1369	GACATGAGACCAACACCAACATGAATACACCACTTACACCACTTACACAGACA	1427
RESULT 12			
US-09-993-966-21			
Sequence 21, Application US/09993966			
Publication No. US20030186232A1			
GENERAL INFORMATION:			
APPLICANT: ROHAN, MICHAEL			
TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NKD PROTEIN,			
FILE REFERENCE: 014024/0280733			
CURRENT APPLICATION NUMBER: US/09/993,966			
CURRENT FILING DATE: 2001-11-27			
PRIOR APPLICATION NUMBER: 60/252,884			
PRIOR FILING DATE: 2000-11-27			
PRIOR APPLICATION NUMBER: 60/291,109			
PRIOR FILING DATE: 2001-05-16			
PRIOR APPLICATION NUMBER: 60/325,571			
PRIOR FILING DATE: 2001-10-01			

```

; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-993-966-21

Alignment Scores:
Pred. No.: 1,59e-90 Length: 590
Score: 1053.00 Matches: 195
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.14% Gaps: 0
DE: 10

US-09-993-966-7 (1-470) x US-09-993-966-21 (1-590)

QY 276 SERPROSERVALALAGLNLYSERGLULEUPROFQAIYTHSERENPOTHTAIGSER 2295
Db 3 TCCCTCTCCGTGGCCAGAAAGTCAGAACTGCCCCCCCGCACTCCCAATCCCATCTCATCT 62
QY 296 ARGSETHIAGLPIRGGLNALAIETHSIETPEQHIARGLYAPROGLINGLYVALAPPRO 315
Db 63 CGCTCCATGACCCGGAACCATTCACATCCACACGAAAGCCCCAGGCGTGAACCG 122
QY 316 AASERPHIAPHLEUAPHTHPRQIIEALYVALSERGLULEUNGINGNARGLEU 335
Db 123 GCCTCTTCCACTTCTTGACACCCCAATCGCCAAAGTCTGAGAGCTCCCAAGCGGCTC 162
QY 336 ARGGLYTHRGINAEPGLYSERYSHIAPHVEVALARGSERPROLYAALAGINGLYLSER 355
Db 183 CGGGGACCCCGAGGACGGAGCAAGCACTTGAGAGTCCCCCAAGGCCAGGCGCAAGAGT 242
QY 356 VALGIVAGLYHIHIVALAARGGLYLAARGAENLYPROPROLEUNGIPROALALE 375
Db 243 GTGGGTGTGGGCGACGTGGCCAGAGGAGGCAAAACACCCCTTGAGAACCGGCATC 302
QY 376 PROALAVASERPROSERALAHISLEUAAIASERPROALALEUENPROSELEUALA 395
Db 303 CCTGGGTGTCCCTCCCTCGCCCACTGGGTGCACGCGGCTCTCTCCCTCCATGCC 362
QY 396 PROLEUNGILYHILYELYSHIILYSHIISARGALALYGLUSERGLINGLYCYVARGLY 415
Db 363 CCCCTCGGGCACAAAGACACAAAGACGAGCCAAAGAGACGACGAGGCTGCCGGGC 422
QY 416 LEUNGINALPROLEUALASERGLYGLYPROVALLEUNGILYARGLUNHISLEURGLULEU 435
Db 423 CTGACGGGACCACTGGCTCCAGTGGCTGTCTCGGGGGGGGAGACCTGCGGAGCTG 482
QY 436 PROALALEUVALIYALTYGLUSERGINAAGIYGLINPROVALGINARGHISGLUNHIS 455
Db 483 CCGGCTTGTGTGTATATAGAGCCAGGCGGGGACGCGGTCCAGAGAATGAGCACAC 542
QY 456 HISHISHISGLUNHISHSIHSIHSITRYHISHSIPHERYGLINTHR 470
Db 543 CACGACCATGAACATCACACCATTAACCACTTACACAGACA 587

RESULT 13
US-09-764-868-152
; Sequence 152, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P1732
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 152
; LENGTH: 1966

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-868-152

Alignment Scores:
Pred. No.: 7.65e-69      Length: 1966
Score: 831.50           Matches: 218
Percent Similarity: 51.74%      Conservative: 49
Best Local Similarity: 42.25%    Mismatches: 142
Query Match: 33.27%           Indels: 107
DB: 9                      Gaps: 19

US-09-993-966-7 (1-470) x US-09-764-868-152 (1-1966)

QY      1 MetGlyLysLeuHisSerIysProAlaAlaValys---LysArgArgLysSerProGlu 19
D      252 ATGGGAAATGTCAGTCGAAGACACGCGCGCGCGCGCAAGCGAAGAGAGCGCGGAA 311
QY      20 -----GlyAspSerPheAlaValSerAla---Ala 28
D      312 GGTTCCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 371
QY      29 TrpAlaArgLysGlyIleGluGluTrpIleGlyArg-----GlnArgCysPro 44
D      372 AGCGCGCGCAAAAGCGCGAGAAAGCGAGCGCGCGCGCGCGCGCAAGCAGAGCTGCC 431
QY      45 GlyGly-----ValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGlyArgSer 62
D      432 AATGGGAGCCCAAGAGAGGCGCTTTCGCG----- 461
QY      63 ThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGluAspAsp 82
D      462 -----GAGGACCAAG 470
QY      83 PheArgLeuGluValAlaLeuProGluLysThrAspGly----- 96
D      471 TGTCCTTCAAGGTCGACTCCCGCTGAGAAAGCTAGAGGCGCGAGACCCGAGCA 530
QY      97 ----LeuGlySerGlyAsp---GluLysLysMetGluArgValSerGluProCysProGly 114
D      531 CTCCTCAGCGAGATGACGAGAGAGGAGCAAAACCGAGGCGCGCGAGACCCGAGC 590
QY      115 SerLysLysGlnLeuLysPheGluGluLeuGlnCysAspValSerMetGluLysAspSer 134
D      591 GGGAGAGCGC---CTCAACATTGACGCACTCCAGTGCATGCTCTCGTGGAGAGAGAC 647
QY      135 ArgGlnGluTrpThrPheThrLeuTyAspPheAspAsnGlnGlyValThrArgGlu 154
D      648 CGCCAGAGTGGACGTTCAAGCTCTATGACTTTGACAACTGCGGGAAGGTCAACAGGAG 707
QY      155 AspIleThrSerLeuLeuHisThrIleTyGluValValAspSerSerValAsnHisSer 174
D      708 GACATGTCAGGCTTCATGACACACATCTATAGGTCGTGATGCTCGGTCAACCACTCC 767
QY      175 ProThrSerSerLysMetLeuArgValLysLeuThrValAlaProAspGlySerGlnSer 194
D      768 TCGGCGCAGCAGCAAGACCTTCCTGCTGTAAGCTTAACCGTCAAGCTTCAAGCAG 827
QY      195 LysArgSerValLeuValAsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThr 214
D      828 AGAAGAGAGGGTCTCTGCTGCGCCAGAGCCGAGAGCCACCGTTGCAAGATGAGAGGT 887
QY      215 LysProThrGluAspLeuArgSerTrpGluLysLysGlnArgAlaProLeuLysPheGln 234
D      888 GAATCTGACAGAGGAGCCAGAGGCTGTCAGCAGAGGTTGTCTGCACTGACAGAGGCC 947
QY      235 GlyAspSerArgLeuGluGlnSerGlyCysTyThrHisCysValAspGluAsnIleGlu 254
D      948 AGTACTAGACCCCGACCGCTGCTGAGAGGGGGCGCCCTTGTGCGTGGACGAGAACGAG 1007
QY      255 ArgArgAsnHisTyLeuAspLeuAlaGlyIleGluAsnTyThrSerGlnPheGlyPro 274
D      1008 CGAGGAACCACTACCTGAGACTCGCGGAGTTTAAAGACTACAGTCAAGTTTGGCCCT 1067

```

```

QY      275 GlySerProSerValAlaGlnLysSerGluLeuProArgThrSerAsn---ProThr 293
D      1068 GGGTCCCTCTCTGTGCAAGCAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1127
QY      294 ArgSerArgSerHisGluProGluAlaIleHisIleProHisArgLysProGln----- 311
D      1128 CGGTCCCGCTCCCAAGAGAGAGAGATACATAGCCGTATACACCCGAGGTCAAGTGTCTG 1187
QY      312 -----GlyValAspProAlaSerPheHisPheLeuAspThrProIleAlaLysValSer 329
D      1188 GTGAAACACGTGTGTGCGAGCTCG-----GAGCTTGTGCTCCGAGCCCTG 1232
QY      330 GlnLeuGlnGlnArgLeuArgGlyThrGlnAspGlySerLysHisPheValArgSerPro 349
D      1233 GACACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1283
QY      350 LysAlaGlnGlyLysSerValGlyValGlyHisValAlaArgGlyAlaArgAsnLysPro 369
D      1284 AAGGCGTCGGGGAAGCGC----- 1301
QY      370 ProLeuGlyProAlaIleProAlaValSerProSer-----AlaHisLeuAlaAlaSer 387
D      1302 -----CTGAGGTGTCCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1352
QY      388 ProAlaLeuLeuProSerLeuAla----- 395
D      1353 CGGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1412
QY      396 ---ProLeuGlnHisLysLysLysHisLysArgAlaLysGluSerGlnGlnGlyCysArg 414
D      1413 CACCCCTAGAGCCCAAGAGGCTACCGCCAAAGAGGAGAGAG-----GGCACTCG 1463
QY      415 GlyLeuGlnAlaProLeuAlaSerGlyGlyProValLeuGluArgGluHis-----Leu 432
D      1464 CCACTCAAGAGCCCAACAGCTCAAGCTGACAGAGT-----GAGCAAGAGTGTG 1514
QY      433 ArgGluLeuProAlaLeuValAlaTyArgLysGlnAlaGlyGlnProValGlnArgHis 452
D      1515 CGGAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1574
QY      453 GluHisHisHisHisHisGluHisHisHisHisHisHisHisHisHisHisHisHis 468
D      1575 GAGGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1622

RESULT 14
US-10-723-860-5220
; Sequence 5220, Application US/10723860
; Publication No. US20040253606A1
GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723, 860
; PRIOR FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5220
; LENGTH: 1959
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1303)..(1320)
; OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-5220

Alignment Scores:

```

Pred. No.: 1,57e-64 Length: 1959
 Score: 786.00 Matches: 215
 Percent Similarity: 47.24% Conservativeness: 50
 Best Local Similarity: 38.32% Mismatch: 152
 Query Match: 31.45% Indels: 145
 DB: 18 Gaps: 20
 US-09-993-966-7 (1-470) x US-10-723-860-5220 (1-1959)
 QY 1 MetGlyValLeuHisSerIysProAlaValCys---LysArgArgIleSerProGlu 19
 Db ATGGGGAAATCGAGTGGAAAGACAGCCGCCGCCGCAAGCGAGAGAGAGCCGGAA 157
 QY 20 GlyAspSerPheAlaValSerAla---AlaTrpAlaArgIleGlyIleGluTrpIle 38
 Db GGGGACAGCTTCCTGGGCTCCGCTACAGCCGCCGCAAGCGCGGAGAGCCGAG 217
 QY 39 GlyArg-----GlnArgCysProGlyGly-----ValSerGlyProArgGln 52
 Db CGCGCGCGCGGACAAAGCAAGAGCTGCCAATGGGAGCCCAAGAGAGGCGCTTCGG 277
 QY 53 LeuArgLeuAlaGlyThrIleGlyArgSerThrArgGluLeuValGlyAspValLeuArg 72
 Db 277 ----- 277
 QY 73 AspThrLeuSerGluGluGluAspPheArgLeuGluValAlaLeuProProGlu 92
 Db 278 -----GAGGACAGTGTCCCTACAGGTGGGACCTCCCGCTGAG 316
 QY 93 LysThrAspGly-----LeuGlySerGlyAsp---GluValCys 104
 Db 317 AAGGAGAGGAGCGCGAGACACCGGAGCAACTCTTACGCGAGATGAGAGAGGAGCA 376
 QY 105 MetGluArgValSerGluProCysProGlySerLysGlnLeuLysPheGluGluLeu 124
 Db 377 GCAACCGCAGAGGCGCGAGAGCCGCGGCGGAGAGCGC---CTCAACATTGACCGCATC 433
 QY 125 GlnCysAspValSerMetGluGluAspSerArgGlnIleTrpThrPheThrLeuTrpAsp 144
 Db 434 CAGTGGATGTCTCGGTGAGAGAGAGACGACCGCAGAGAGTGTGACGTTACGCTCTATGAC 493
 QY 145 PheAspAsnAsnGlyLysValThrArgGluAspIleThrSerLeuLeuHisThrIleTrp 164
 Db 494 TTGGACACTGCGGAGAGGTTCACCGAGAGAGACATGTCAGCTTCATGACACACATTAT 553
 QY 165 GluValValAspSerSerValAsnHisSerProThrSerSerLysMetLeuArgValLys 184
 Db 554 GAGGTGTGGATGCTCGTCAACCACTCTCGGCGAGAGAACCTCCGTGTGAG 613
 QY 185 LeuThrValAlaProAspGlySerGlnSerLysArgSerValLeuValAsnGlnAlaAsp 204
 Db 614 CTAAACGCTCAGCCCTGAGCCCTCAGCAAGAGAGAGGAGTCTCTCTGCGCCAGGAC 673
 QY 205 LeuGlnSerLysArgProArgAlaGluThrLysProThrGluAspLeuAspSerTrpGlu 224
 Db 674 CGGAGACCCACCCGTTGACAGATGAGAGGTAACTGGCAGAGAGAGCCAGGAGCGTAC 733
 QY 225 LysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGluGlnSerGlyCys 244
 Db 734 AGAGAGTTGTCTCCACAGTCACAGAGGCCAGACTGACCCCGAGCCCTCTCGAGCGG 733
 QY 245 TyrHisHisCysValAspGluAsnIleGluArgArgAsnHisGlyLeuAspLeuAlaGly 264
 Db 794 GGGCCCTACTGCTGCGAGAACACCGAGCCGAGAAACCATACCTGAGCTCGCGGG 853
 QY 265 IleGluAsnTrpThrSerGlnPheGlyProGlySerProSerValAlaGlnLysSerGlu 284
 Db 854 ATTGAATACTACAGTCCAGATTGCGGCTGAGCTTCACTGTGCGAGAGAGAGCTCC 913
 QY 285 LeuPro----- 286
 Db 914 GCTCCACAGACACAGTGGGAGCAAGGCTAGAGAGTGGCCTTTGACAGAGACTGTGG 973

QY 286 ----- 286
 Db 974 AGCCAGGACAGTACCCACAGTGGCCAGGCCCTTCCCTTCAAGGACTGTGCGCTGTGA 1033
 QY 286 ----- 286
 Db 1034 CTGCAGACTTGCTTAACAGACTGGCTCAGAGTCCCTCTGTGCAAGCAAGAGAGAGC 1093
 QY 287 -----ProArgThrSerAsnProThrArgSerArgSerHisGluProGluAla 302
 Db 1094 CCCAGGACAGGAGCTTGCACTCCAGGCCCGG---TCCGCTTCCAGAGAGAGATGACA 1149
 QY 303 IleHisIleProHisArgGlySerProGln-----GlyValAspProAlaSerPhe 318
 Db 1150 CATGCCGTACACACCGCAGGTCAAGTGTGTGTGAGACACGTCGTGCGAGCTCG--- 1206
 QY 319 HisPheLeuAspThrProIleAlaLysValSerGluLeuGlnGlnArgLeuArgGlyThr 338
 Db 1207 -----GAGCTGTCTGCCCGGCGCTTGACACGAGCCCGCGGAGAGGCGCG 1254
 QY 339 GlnAspGlySerLysPheValArgSerProLysAlaGlnGlyLysSerValGlyVal 358
 Db 1255 GAG-----AAGCAGTTCCTCAAGTCCCGCAAGGAGCTCGGAGAGCGCGTGGGNN 1305
 QY 359 GlyHisValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaVal 378
 Db 1306 NNNNNNNNNNNNNNTCCGGGAAA-----GCTTCAGCTACTACCTGCGCGCTC 1356
 QY 379 SerPro-----SerAlaHisLeuAlaHisSerProAlaLeuLeuPro 392
 Db 1357 CTGCGCCCGGAGCCCTCAGACAGCGCCACCACTCTCCGAGCCCGCCAGC----- 1407
 QY 393 SerLeuAlaProLeuGlnHisLysLysValHisIleArgAlaLysGluSerGlnGlnGly 412
 Db 1408 -----CAACCTTACGGCCCAAGACGGTACCGCCAAAGGACAGAGAG-----GCC 1452
 QY 413 CysArgGlyLeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlyAlaGluHis--- 431
 Db 1453 CACTCGCACTCAAGGCCCCCAACAGCTCAGCTGCCACACTG-----GAGCACAGAG 1503
 QY 432 ---LeuArgGluLeuProAlaLeuValValTyrGluSerGlnAlaGluGlnProValGln 450
 Db 1504 GTGTGCGGAGACTTGTCCGCCAGCCAGCCAGAGAGAGAGGCTTACCGCTCAGATGATCAG 1563
 QY 451 ArgHisGluHisIleHisIleHis-----HisGluHisIleHisIleHisIleHisPhe 467
 Db 1564 CGGACAGAGACACACACACACACACACAGAGACACACACACACACACACACACTTC 1623
 QY 468 Tyr 468
 Db 1624 CAC 1626
 RESULT 15
 US-09-918-995-8686
 ; Sequence 8686, Application US/09918995
 ; Publication No. US20030073623A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hyeq, Inc.
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
 ; FROM VARIOUS CDNA LIBRARIES
 ; FILE REFERENCE: 20411-756
 ; CURRENT APPLICATION NUMBER: US/09/918, 995
 ; CURRENT FILING DATE: 2001-07-30
 ; PRIOR APPLICATION NUMBER: US/09/235, 076
 ; PRIOR FILING DATE: 1999-01-20
 ; NUMBER OF SEQ ID NOS: 38054
 ; SOFTWARE: PacBioSBO for Windows Version 3.0
 ; SEQ ID NO 8686
 ; LENGTH: 427
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature

LOCATION: (1)...(427)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-918-995-8686

Alignment Scores:

Pred. No.:	6.37e-61	Length:	427
Score:	739.00	Matches:	140
Percent Similarity:	99.29%	Conservative:	0
Best Local Similarity:	99.29%	Mismatches:	1
Query Match:	29.57%	Indels:	0
DB:	10	Gaps:	0

US-09-993-966-7 (1-470) x US-09-918-995-8686 (1-427)

```

QY      285  LeuProProArgThrSerAnProThrArgSerArgSerHisGluProGluAlaIleHis 304
      |||
      4  CTGCCCCCGCACCTCCCAATCCCACTCGATCTGCCATGAGCGGAAGCCATCCAC 63
      |||
QY      305  IleProHisArgLysProGlnGlyValAspProAlaSerPheHisPheLeuAspThrPro 324
      |||
      64  ATCCCAACCGAAGAGCCCAAGGCGTGACCCGCGCTCTTCCACTTCTTGACACCCCA 123
      |||
QY      325  IleAlaIysValSerGluLeuGlnIleArgLeuArgGlyThrGlnAspGlySerLysHis 344
      |||
      124  ATGCGCAAGGCTCAGAGCTCCAGCAAGGCTCCGGGGCACCCGAGCGGAGCAAGCAC 183
      |||
QY      345  PheValArgSerProLysAlaGlnGlyLysSerValGlyValHisValAlaArgGly 364
      |||
      184  TTGTGTAGGTCCCGCAAGGCCAGAGGCAAGAGTGTGGGTGTGGGCCAGTGGCCAGAGGG 243
      |||
QY      365  AlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerProSerAlaHisLeu 384
      |||
      244  GCAAGAAACAAGCCCTCTGGGACCCGCCATCCCTCGGGTCTCCCTCCGCCCACTG 303
      |||
QY      385  AlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLysLysHis 404
      |||
      304  GCTGCCAGCCCGGCTCTCTCCCTCTAGCCCTCGGGCAAGAAGACCAAGCAC 363
      |||
QY      405  ArgAlaIysGluSerGlnGlnGlyCysArgGlyLeuGlnIleAlaProLeuAlaSerGlyGly 424
      |||
      364  CGAGCCAAAGAGAGCAAGAGGCTGCGGGGCTTGCAAGCACCACTGGCTCAAGTGGC 423
      |||
QY      425  Pro 425
      |||
      424  CCT 426
  
```

Search completed: December 30, 2004, 09:35:19
 Job time : 5904 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 29, 2004, 21:21:19 ; Search time 848 Seconds

(without alignments)
2909.467 Million cell updates/sec

Title: US-09-993-966-7

Perfect score: 2499
Sequence: 1 MGKLSKRAVCKRRESREG.....RHHNNHHNNHHNNHYQT 470

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=fprime+ p2n.model -DEV=xlp
-O=/cgn2.1/USPTO.spool.p/US0993966/runat.29122004.130757.8986/app.query.fasta_1.647
-DB=N Geneseq.23Sep04 -QFMT=fastsp -SUFFIX=ring -MINMATCH=0.1 -LOOPTCL=0
-LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR.SCOR=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USBR=US0993966_@CGN.1.1.885_@runat.29122004.130757.8986 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_23Sep04:*

1: geneseqn1980s:*\n2: geneseqn1990s:*\n3: geneseqn2000s:*\n4: geneseqn2001as:*\n5: geneseqn2001bs:*\n6: geneseqn2002as:*\n7: geneseqn2002bs:*\n8: geneseqn2003as:*\n9: geneseqn2003bs:*\n10: geneseqn2003cs:*\n11: geneseqn2003ds:*\n12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2499	100.0	1859	10	AA151511
2	2493	99.8	1438	3	AA63925
3	2192.5	87.7	1416	3	AAH28343
4	2174.5	87.0	1731	3	AA63924
5	2171.5	86.9	1401	10	AA151512
6	1067	42.7	2379	4	AA103214

7	1067	42.7	2379	4	AA103216	AA103216 Human rep
8	1053	42.1	590	10	AA151525	AA151525 Human Nkd
9	842	33.7	1743	5	AA878752	AA878752 DNA encod
10	842	33.7	1743	5	AA892081	AA892081 DNA encod
11	842	33.7	1743	10	ADCC1966	ADCC1966 Human nov
12	831.5	33.3	1966	4	AA827117	AA827117 CDNA encod
13	831.5	33.3	1966	10	ADBB3295	ADBB3295 Human cdn
14	798.5	32.0	1561	5	AA892082	AA892082 DNA encod
15	786	31.5	1959	12	ADQ22400	ADQ22400 Human sof
16	783	31.3	1885	3	AA63926	AA63926 DNA encod
17	780.5	31.2	1285	3	ADFF82455	ADFF82455 Leukaemia
18	739	29.6	427	9	ACH21474	ACH21474 Human adu
19	583.5	23.3	1307	3	AA63927	AA63927 DNA encod
20	378.5	15.1	1817	6	ABK34945	ABK34945 Human cdn
21	271	10.8	639	8	ACD05637	ACD05637 CDNA encod
22	246	9.8	1950	6	ABQ73820	ABQ73820 Human col
23	239	9.6	598	12	ACH74795	ACH74795 Human gen
24	238	9.5	148	10	AA151522	AA151522 Human Nkd
25	237	9.5	128	10	AA151524	AA151524 Human Nkd
26	233	9.3	133	12	ACH88495	ACH88495 Human gen
27	233	9.3	134	10	AA151518	AA151518 Human Nkd
28	215.5	8.6	2691	4	ABL24299	ABL24299 Drosophill
29	215.5	8.6	4954	3	AA63923	AA63923 DNA encod
30	209	8.4	2723	4	AAK75944	AAK75944 Human imm
31	186	7.4	107	10	AA151520	AA151520 Human Nkd
32	182	7.3	3486	4	ABL27205	ABL27205 Drosophill
33	175	7.0	96	10	AA151521	AA151521 Human Nkd
34	172	6.9	6749	12	ADQ24048	ADQ24048 Human sof
35	170.5	6.8	2740	3	AACT5288	AACT5288 Human ORF
36	164	6.6	8104	4	ABL24298	ABL24298 Drosophill
37	164	6.6	51855	10	ACR04818	ACR04818 Melitiaz
38	161.5	6.5	554	12	ACH74420	ACH74420 Human gen
39	159	6.4	9807	4	ABL06949	ABL06949 Drosophill
40	159	6.4	23407	4	ABL06948	ABL06948 Drosophill
41	158.5	6.3	2766	4	ABL07259	ABL07259 Drosophill
42	156.5	6.3	1633	10	ADCG7022	ADCG7022 Human GPC
43	156.5	6.3	2981	5	AA570530	AA570530 DNA encod
44	156.5	6.3	2981	10	ADBO9862	ADBO9862 Novel DNA
45	156.5	6.3	6714	4	ABL16321	ABL16321 Drosophill

ALIGNMENTS

RESULT 1					
AA151511	AA151511 standard; DNA; 1859 BP.				
AC	AA151511;				
DT	24-APR-2003 (first entry)				
DE	Human Nkd (hNkd) protein coding sequence.				
XX	Human, gene; ds; gene therapy; Nkd; hNkd; Wnt signalling pathway; cancer;				
KM	colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer.				
XX	Homo sapiens.				
OS					
XX					
FH	Key	Location/Qualifiers			
FT	CDS	224..1636			
FT		/tag= a			
FT		/product= "Human Nkd protein"			
XX					
PN	MO200292832-A2.				
PD	21-NOV-2002.				
XX					
PF	27-NOV-2001; 2001WO-US044092.				
XX					
PR	27-NOV-2000; 2000US-0252884P.				
PR	16-MAY-2001; 2001US-0291109P.				
PR	01-OCT-2001; 2001US-0325571P.				
XX					

FT XX /product= "naked cuticle polypeptide"

PN MO200049034-A1.

XX 24-AUG-2000.

PD 17-FEB-2000; 2000MO-US004188.

XX 17-FEB-1999; 99US-0120646P.

PR (STRD) UNIV LEIAND STANFORD JUNIOR.

PA Scott M, Zeng W, Wharton K;

XX MPI, 2000-571967/53.

DR P-PSDB; AAB08216.

XX An isolated nucleic acid molecule useful for analyzing (genetic

PT predileposition to) s disease state and for therapeutic purposes e.g.

PT treatment of cancer comprises a sequence encoding a naked cuticle

PT protein.

XX Claim 3; Page 46-48; 58pp; English.

PS The present sequence encodes a protein related to the Drosophila Nkd

XX (naked cuticle) polypeptide. In Drosophila, Nkd is a segment-polarity

CC gene whose expression is induced by Wnt signalling. The Nkd polypeptide

CC acts to antagonize Wnt signalling. Nkd may link ion fluxes to the

CC regulation of Wnt signal potency, duration or distribution. The Nkd

CC polynucleotides can be used for identifying homologous or related

CC proteins, to modulate the expression or function of Nkd polypeptides, and

CC in studying associated physiological pathways. Nkd polynucleotides can

CC also be used in gene therapy to treat disorders associated with Nkd

CC defects. They may also be used for therapeutic purposes e.g. treatment of

CC cancer

XX Sequence 1438 BP; 319 A; 478 C; 439 G; 202 T; 0 U; 0 Other;

XX

Alignment Scores:

Pred. No.:	5.08e-147	Length:	1438
Score:	2493.00	Matches:	469
Percent Similarity:	99.79%	Conservative:	0
Best Local Similarity:	99.79%	Mismatches:	1
Query Match:	99.76%	Indels:	0
DB:	3	Gaps:	0

US-09-993-966-7 (1-470) x AAA63925 (1-1438)

QY 1 MerGlyLylLeuHissertlypProAlaAlaValCylYsArGArGluSerProGluGly 20

DB 8 ATGGGAAACTTCACTCCAAAGCCGCGCGTGTGCAAGCGAGGAGACCCGGAAAGGT 67

QY 21 AspSerPheAlaValSerAlaAlaATPPAlaArgLysGlyTleGluGluTPIleGlyArg 40

DB 68 GACAGCTTCCCTGACCTGAGCTGAGCTCGAAAGGCAATCGAGAGTGAATCGAGAGA 127

QY 41 GlnArgCyseProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGly 60

DB 128 CACCGCTGCGCGGCGGTGTCTCGGAGCCCGACAGCTGTGGCGGACCAATAGGC 187

QY 61 ArgSerThrArgLysLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80

DB 188 CGAAGACCCCGGAGCTCTGTGGCGAGAGTGTAGAGACACGCTCACCGAGAAAGAGAG 247

QY 81 AspAspPheArgLeuGluValAlaLeuProProGluLysThrAspGlyLeuGlySerGly 100

DB 248 GACGACTTTTGGCTGAAAGTGGCCCTCTCTTGAAGAGACTGACGGGCTGGGACCGGA 307

QY 101 AspGluLysLysMetGluArgValSerGluProCyseProGlySerLysGlnLeuLys 120

DB 308 GATGAGAAAGATGAGAGAGTGAAGGAAACCTTGCCCAAGAGGAGAGCTGAAG 367

QY 121 PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGlnGluTrrPhe 140

DB 368 TTGAAGAGCTCCAGTCCGAGCTGTCTCATGAGAGAGACGCGGAGAGTGAAGCTTTC 427

QY 141 ThrLeuTyzAspPheAspAsnAenGlyLysValThrArgGluAspIleThrSerLeuLeu 160

DB 428 ACCCTGTATGACTTTGACAAACAGGAGAGTCAACCGAGAGAGACATCACAGCTTCTG 487

QY 161 HisThrIleTyrgluValValAspSerSerValAsnHisSerProThrSerSerLysMet 180

DB 488 CACACATCTATAGAGTGTGAGCTCTCTGTCAACCACTCCCAATCATCTACAGAAATG 547

QY 181 LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysAspSerValLeuVal 200

DB 548 CTGGGAGTAAAGTCACTGAGCTGCGCCCGATGGAGCCAGACGAAAGAGAGCTTCTTTC 607

QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrArgLysAspLeu 220

DB 608 AATCAGGCTGACTGCGAGAGCGGAGAGCCCGAGAGAGACCAAGCCACTGAGAGACTG 667

QY 221 ArgSerTrpGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240

DB 668 CGGAGCTGGGAGAAAGACGACGAGCCCGCTCAGGTTCCAGGGGTGACACCGCTTGAG 727

QY 241 GlnSerGlyCyseThrHisCysValAspGluAsnIleGluArgArgAsnHisTyrrLeu 260

DB 728 CAGCTGTGCTGTACCACTTCGTAATGATGAGAACATCGAGAGAGAAACCACTACTTA 787

QY 261 AspLeuAlaGlyTleGluLeuThrSerGlnPheGlyProGlySerProSerValAla 280

DB 788 GATCTGCGCGGATAGAAACCTACGTCCCAATTTGGGCTGAGCTCTTCCGTGGCC 847

QY 281 GlnLysSerGluLeuProProArgThrSerAspProThrArgSerArgSerHisGluPro 300

DB 848 CAGAAATGAGAAATGCGCCCGCCCGCTCAATCCACTGATCTGCTCCCTGAGCCG 907

QY 301 GlnAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPhe 320

DB 908 GAAGCCATCCACATCCCAACCGAAAGCCCAAGGGGTGAGCCGGCTCTTCCACTTC 967

QY 321 LeuAspThrProIleAlaLysValSerGluLeuGlnArgLeuArgGlyThrGlnAsp 340

DB 968 CTTGACACCCCAATCCGAAAGTCTCAGAGTCCAGCAACGGTCCGGGGCACTCAGAGAC 1027

QY 341 GlySerLysHisPheValArgSerProLysValGlnGlyLysSerValGlyValHis 360

DB 1028 GGGAGCAAGCACTTGTGAGTCCCGCAAGGCCAGGGCAAGAGTGTGGGTGGCCAC 1087

QY 361 ValAlaArgGlyValArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380

DB 1088 GTGGCCAGAGGGGCAAGAAACAAAGCCCTCTGGAGCCCGCAATCCCTGGGTGTCCCC 1147

QY 381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400

DB 1148 TCGGCCCACTGGCTGCGAGCCCGGCTCTCTCCCTTCTTACCCCTCGGGGCAAG 1207

QY 401 LysHisLysHisArgAlaLysGlnSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420

DB 1208 AAGCACAAGACACGAGCAAGAGAGACAGAGGAGGCTGCGGGGCTGCGAGGACCACTG 1267

QY 421 AlaSerGlyGlyProValLeuGlyArgGlnHisLeuArgGluLeuProAlaLeuValVal 440

DB 1268 GCTTCAGTGGGCTCTGTCTGGGGCGAGACACTCGGAGAGTGGCCGCTTGTGGTGTG 1327

QY 441 TyrGlnSerGlnAlaGlyGlnProValGlnArgHisGlnHisHisHisHisHisHis 460

DB 1328 TATGAGAGCCAGGCGGGGAGCGGCTCCAGAGCAATGAGACCAACCAACCAATGAACAT 1387

QY 461 HisHisHisTyrrHisHisPheTyrglnThr 470

DB 1388 CACCAACATTAACCACTTCTACAGACA 1417

RESULT 3

AAH28343

ID	AAH28343 standard; DNA, 1416 BP.
XX	
AC	AAH28343;
XX	
DT	05-SEP-2001 (first entry)
XX	
DE	Nucleotide sequence of a murine nmkd protein.
XX	
KW	Dishevelled-associated protein 1A; DAPIA; nmkd; dishevelled protein;
KV	Wnt signalling; colon cancer; cancer; ss.
XX	
OS	Mus sp.
XX	
FH	Key Location/Qualifiers
FT	1..1416
CDS	/*tag= a
FT	/product= "nmkd"
XX	
FN	MO200144279-A2.
XX	
PD	21-JUN-2001.
XX	
PE	05-DEC-2000; 2000MO-US032986.
XX	
PR	17-DEC-1999; 99US-0172434P.
XX	
PA	(CHIR) CHIRON CORP.
PI	Yan D, Williams LT;
DR	WPI, 2001-425440/45.
XX	P-PsDB; AAB84645.
PT	Novel mammalian dishevelled-associated proteins, nmkd and DAPIA, useful
PT	for inhibiting Wnt signaling in mammalian cells, and thus for treating
PT	colon cancer.
PS	Claim 2; Fig 1; 67pp; English.
XX	
CC	The present sequence encodes a murine nmkd protein. The specification
CC	describes nmkd and dishevelled-associated protein 1A (DAPIA) proteins,
CC	that interact with mammalian dishevelled protein. nmkd is useful for
CC	inhibiting Wnt signaling in mammalian cells, and thus for treating colon
CC	cancer. nmkd are also useful for activating the JNK pathway. nmkd and
CC	DAPIA are also useful for screening drugs that are useful for treating
CC	cancer
SQ	Sequence 1416 BP, 339 A, 434 C, 419 G, 224 T, 0 U, 0 Other;
Alignment Scores:	
Pred. No.:	3,19e-128 Length: 1416
Score:	2192.50 Matches: 410
Percent Similarity:	91.06% Conservative: 18
Best Local Similarity:	87.23% Mismatches: 41
Query Match:	87.74% Indels: 1
DB:	Gaps: 1
US-09-993-966-7 (1-470) x AAH28343 (1-1416)	
OY	1 MetGlyLVLeuHisSerIysProAlaAlaValCysLysArgArgGluSerProGluGly 20
Db	1 ATGGGGAAACTTCACCTCGAAGCCGGCCGCCTGTGCAACGCGAGGAGACC CGAAGGT 60
OY	21 AppSerPheAlaValSerIalaIatTpaIaaTgLyegLIleGlucIuTrpIIleGlyarg 40
Db	61 GACAGCTTTTGTAAGCGCTGCTTGGGAAGGAAAGCATCGAGATGGATCGGAGAG 120
OY	41 GlnaRcYsPrCoGIyValSerGIyProArGIuLeuArgHeuAlagIyThrIIlegIy 60
Db	121 CAGCGCTGTCCAGGCAAGCGCTTCAGAGACCCCGTCACCTAGAATTGGCAGGCACTGTGGT 180
OY	61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrIeuSerGIuGluGlu 80

Dd	181	CGAGGCACTGGGGAACCTCGTGGGGAACACTTCTAGAGAGGCTCTCGGTGAGAGAGACGAG	240
Qy	81	AspAspPheArgLeuGluValAlaLeuProProGluIuysThrAspGlyLeuGlySerGly	100
Dd	241	GACGACCTTCCCTCAGAAAGTGGCCCTCGCGCTGAGAAATGACACACTTAAAGTAAGTGGGA	300
Qy	101	AspGluIuysbMetGluArgValSerGluProCysProGlySerGlyblybGluIuys	120
Dd	301	GATAGAGAGAAATGAGAGAGACTGAGGGAACCTTGGCCAGGCTCTCAAGAGAGACTTAAG	360
Qy	121	PheGluGluLeuGlnCysAspValSerMetGluIuysPheArgGlnGluIuysPThrPhe	140
Dd	361	TTTGAAGAGCTCACTGTGATGTCTCTGTGAGAGAGACAGCCGGCAAGATGGACTTTC	420
Qy	141	ThrLeuTyArgPheAspAsnAsnGlyValValThrArgGluAspIleThrSerLeuLeu	160
Dd	421	ACTCTATATGACTTCGACCAACATGAGGAAGTGAACCGGTGAGGACATTACAGCTTCTG	480
Qy	161	HisThrIleTyArgGluValAlaSerSerSerValAsnHisSerProThrSerSerIleMet	180
Dd	481	CATACCATCTATGAAAGGTGTGACTCTCTGTGTGAACCATTCCTCCACATCAAGCAACACA	540
Qy	181	LeuArgValIuysLeuThrValAlaProAspGlySerGlnSerIuysAspSerValLeuVal	200
Dd	541	CTGGCGGTGAAGCTCACCGGTGCTCTGACGGAGCCAGAGTAAGAGAGAGCTCTTTC	600
Qy	201	AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIuysProThrGluAspLeu	220
Dd	601	AAACCATACCGACTGCGAGAGACACAAAGGCCCGGACAGACACAAACCCGCTGAGAGACTG	660
Qy	221	ArgSerTrpGluIuysblybGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu	240
Dd	661	CGTGAGCTGGGAGGAAGACGCGAGCCCACTCAGGTTCAGAGGTGACACCACTGGAG	720
Qy	241	GlnSerGlyCysTyTrpHisCysValAspGluAsnIleGluArgArgAsnHisTyLeu	260
Dd	721	CAGCCAGACTCTACCAACCATTTGGTGTGATGAGAACATTGAGAGAGAAACCACTACTTA	780
Qy	261	AspLeuAlaGlyIleGluAsnTyThrSerGlnPheGlyProGlySerProSerValAla	280
Dd	781	GACCTGGGGGGAGTAAAGAACTACAGCTTCAGTTTGGACCGGATCTCCCTTGGGTGGC	840
Qy	281	GlnIuysSerGluLeuProProArgTrpSerAsnProThrArgSerArgSerHisGluPro	300
Dd	841	CAGAAGTGAGAGCTCCCTCGAATCTCCAAACCCCACTGGCTCTCCCTCCACAGAACCA	900
Qy	301	GluAlaIleHisIleProHisArgIuysProGlnGlyValAspProAlaSerPheHisPhe	320
Dd	901	GAAAGTGGCCCAATCCCAACCGGAGCCCAAGGTGTGAACCAAGGCTCTTTCACCTC	960
Qy	321	LeuAspThrProIleAlaIuysValSerGluIuysGlnGlnArgLeuArgGlyThrGlnAsp	340
Dd	961	CTTGAACCCCACTTTGGCAAGCATCAAGGCTCCAGCAAGGCTCGGGGAGCTCAAGAT	1020
Qy	341	GlySerIuysHisPheValArgSerProIuysAlaGlnGlyIuysSerValGlyValGlyHis	360
Dd	1021	GGGAGCAAGCACTTTGTGAGGTCTCCCAAGGCTCCAGGGCAAGAACATGGGTATGGGCAC	1080
Qy	361	ValAlaArgGlyAlaArgAsnIuysProProLeuGlyProAlaIleProAlaValSerPro	380
Dd	1081	GGGGCCAAAGGTGAAGAAAGCAACCTCCACTGTGTACCAACCACTACTGTCTTCCCC	1140
Qy	381	SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisIuys	400
Dd	1141	TCTGCGCATCTGGCCACCAAGCCGCTTCTCCCAACCTTGGGACACCTTGGGGCACAAG	1200
Qy	401	IuysHisIuysHisArgAlaIuysGluSerGlnGlnIuysCysArgGlyLeuGlnAlaProLeu	420
Dd	1201	AAACCAAGACATCGAGCCCAAGAGACGAGCGAGCTGCGGGGCTGACAGGGCCCCCTG	1260
Qy	421	AlaSerGlyGly---ProValLeuGlyAlaArgHisLeuArgGluIuysProAlaLeuVal	439
Dd	1261	GCTGGAGAGGCTCACCGTCAATGGGGCGGAGACAGAGAGAGAGGCTGCTGGCGGTGGT	1320

OY		440 VALTYRGUSEGALAGLYGNPRVGLNARQHIHSHISHISHLIN 459 DB	1321 GELTGCGAAGCCACGGCTGTGGCAGGCCGTGCAGAACAACATCACCAACGA 138
OY		460 HHSIHSHISTYRHHSHISPHERYGJLN 469 DB	1381 CATCACACCATTATTCACCATTTCTTATCG 1410
RESULT 4			
ID	AAA63924	standard; DNA; 1731 BP.	
AA	AAA63924;		
XX	AAAA3924;		
DT	04-DEC-2000	(first entry)	
DE	DNA encoding protein related to Drosophila naked cuticle polypeptide.		
XX	Nkd gene; naked cuticle gene; segment-polarity gene; Wnt signalling;		
KW	gene therapy; Nkd defect; cancer; ss.		
OS	Mus musculus.		
FT	Key	Location/Qualifiers	
CDS	. . .	140..1555	
FT	*tag= a	/product= "naked cuticle polypeptide"	
PN	WO200049034-AI.		
PD	24-AUG-2000.		
XX	17-FEB-2000; 200OMO-US004188.		
PE	17-FEB-1999; 99US-0120646P.		
PR	(STRD) UNIV LELAND STANFORD JUNIOR.		
PA	Scott M, Zeng W, Wharton K;		
PI	WP1: 2000-571967/53.		
DR	P-FSDB; AAB08215.		
XK	An isolated nucleic acid molecule useful for analyzing (genetic predisposition to) a disease state and for therapeutic purposes e.g. treatment of cancer comprises a sequence encoding a naked cuticle protein.		
PT			
PS	Claim 3; Page 43-45; 58pp; English.		
XX	The present sequence encodes a protein related to the Drosophila Nkd (naked cuticle) polypeptide. In Drosophila, Nkd is a segment-polarity gene whose expression is induced by Wnt signalling. The Nkd polypeptide acts as antagonist Wnt signalling. Nkd may link ion fluxes to the regulation of Wnt signal potency, duration or distribution. The Nkd polynucleotides can be used for identifying homologous or related proteins, to modulate the expression or function of Nkd polypeptides, and CC in studying associated physiological pathways. Nkd polynucleotides can also be used in gene therapy to treat disorders associated with Nkd defects. They may also be used for therapeutic purposes e.g. treatment of cancer.		
SQ	Sequence 1731 BP; 414 A; 521 C; 504 G; 292 T; 0 U; 0 Other;		
Alignment Scores:			
Pred. No.:	5.37e-127	Length:	1731
Score:	2174.50	Matches:	407
Percent Similarity:	90.64%	Conservative:	19
Best Local Similarity:	86.60%	Mismatches:	43
Query Match:	87.01%	Indels:	1
GB:	3	Gaps:	1

US-09-993-966-7 (1-470) x AAA63924 (1-1731)

QY	1	MecGlyLysLeuHISerLysProAlaAlaValCylsYsAArgArgLysSerProGluGly	20
Db	140	ATGGGGAATCTTCACTCGAAGCCGGCCGCGCTGTGCAAGCCAGGAGAGCCGGAAAGT	199
QY	21	AspSerPheAlaValSerAlaAlaThrAlaArgLysGlyIleGluGluTrpIleGlyArg	40
Db	200	GACAGCTTGTGTGAAGCGCTGCTTGGGCAAGAAAGGCAATCGAGATGTGATCCGGAGG	259
QY	41	GlnArgCysProGluGlyValSerGlyProArgGlnLeuAArgLeuAlaGlyThrIleGly	60
Db	260	CAGCCCTTCCAGGAGCAGCGCTCTCAGAGACCCCTCAGCTGAAATTTGGCAGGCACTTGT	319
QY	61	ArgSerThrArgGlnLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu	80
Db	320	CGAGGCACTCGGGAACCTGTGGGTGACACTTCTAGAGAGGCTCTCCGTGAGAGAGGACAG	379
QY	81	AspAspPheAlaGluGluValAlaLeuProProGluGlyThrAspGlyLeuGlySerGly	100
Db	380	GACACATTCCTCCCTAGAAAGTGCCCTGCGGCTGAGAAAGATCGACAGCTTGTAATGGA	439
QY	101	AspGluLysLysMetGluArgValSerGlnProCysProGlySerLysLysGluLeuLys	120
Db	440	GATGAGAGAAAGATGAGAGACGTAGCCGAACTTGCCAGCGCTCCAGAAACAGCTCAAG	499
QY	121	PheGluGlnLeuGlnCysAspValSerMetGluLysAspSerArgGlnGluTrpThrPhe	140
Db	500	TTTGAAGAGCTACAGTGTGATGTCTGTGTGAGAGAGAGCAGCCGGAAGAATGGACTTTC	559
QY	141	ThrLeuTyraaPheAspAsnAsnGlyLysValThrArgGluAspIleThrSerLeuLeu	160
Db	560	ACTCATATATGACTTCCACAAATGGAATGACCCGTGAGGACATTACACAGCTTGCTG	619
QY	161	HisThrIleTyrgluValValAspSerSerValAsnHisSerProThrSerSerLysMet	180
Db	620	CATACCATCTATGAAGGTGTGACTCTCTGTGTAACATTCCTCCACATCAAGCAAGACA	679
QY	181	LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysArgSerValLeuVal	200
Db	680	CTGGGGGTGAAGCTCACCGTGGCTCTGACGGAGCCAGAGCTAAGAGAGGCTCTTTC	739
QY	201	AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeu	220
Db	740	AACCATACCGATCTGCAGAGACACAAGGCCCGAGACAGACACAAACCGGTGAGAGCTG	799
QY	221	ArgSerTrpGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu	240
Db	800	CGTGAGCTGGGAGAAAGACACCGAGCCCACTCAGGTCCAGGGGTGACAGCACCTGGAG	859
QY	241	GlnSerGlyCysTyrglnHisCysValAspGluAsnIleGluAspArgAsnHisTyrlleu	260
Db	860	CAGCCAGACTCTCTACACCATTTGGTGTGATGAGAACATTGAGAGAGAAACACTTACTTA	919
QY	261	AspLeuAlaGlyIleGluAsnTyrlThrSerGlnPheGlyProGlySerProSerValAla	280
Db	920	GACCTGGGGGAGATAGAGAACTACAGCTCAGTTTGACCGGGATCCCTTCGGTGGCC	979
QY	281	GlnLysSerGluLeuProProArgTrpThrAsnProThrArgSerArgSerHisGluPro	300
Db	980	CAGAAAGTCAGAGCTGCCCTCCGATCTCCAAACCCCACTGGCTCTGCCCAAGAGCA	1039
QY	301	GlnAlaIleHisIleProHisArgLysLysProGlnGlyValAspProAlaSerPheHisPhe	320
Db	1040	GAACTTGCACATCCACACCGGAGGCCCAAGATGTGAACCCAGGCTCTTCACTTC	1099
QY	321	LeuAspThrProIleAlaLysValSerGluLeuGlnLysArgLeuArgGlyThrGlnAsp	340
Db	1100	CTTGACACCCCATTTGGCAAGCATCAGAGTCCAGCAAGCGCTCCGGGGGCACTCAAGAT	1159
QY	341	GlySerLysHisPheValaArgSerProLysAlaGlnGlyLysSerValGlyValGlyHis	360
Db	1160	GGGAGCAAGACATTTGTGAGGTCTCCCAAGGCCCAAGGCAAGAACATGTGATGGGCAC	1219

QY 361 ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
 |||||
 Db 1220 GGGGCCAGAGGTGCAGAGAGACCTCCACTGGTACCCACCACCTACTGTCCTCCCC 1279
 |||||
 QY 381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyValLys 400
 |||||
 Db 1280 TCTCCCATCTGGCCACGAGCCCTTCTCCCAACCTGGCACCCCTGGGACACAG 1339
 |||||
 QY 401 LysHisLysHisAlaGlyAlaLysGlySerGlnGlnLysGlySerGlyLeuGlnAlaProLeu 420
 |||||
 Db 1340 AAACACAGACATCGAGCCAGAGAGACGCGAGCTGCCGGGCTTCGAGGCCCTG 1399
 |||||
 QY 421 AlSerGlyGly--ProValLeuGlyAlaArgLysLeuArgGlyLeuLeuProAlaLeuVal 439
 |||||
 Db 1400 GCTGCAAGAGGCTCCACCGTCATGGGCGGAGCAGGTGAGGAGCTGCTGCTGCTGATG 1459
 |||||
 QY 440 ValTyrGlySerGlnAlaGlyGlnProValGlnArgHisGlyHisLysHisLysGly 459
 |||||
 Db 1460 GTGTACGAGAGCCAGGCTAGGCAAGCCGTCAGAGACACGAAACCATCACACGAA 1519
 |||||
 QY 460 HisHisHisHisLysTyrHisHisLysPheTyrGln 469
 |||||
 Db 1520 CATCACCACTATATACACTTCTATCAG 1549
 |||||

RESULT 5
AAL51512

ID AAL51512 standard; DNA, 1401 BP.

AC AAL51512;

DT 24-APR-2003 (first entry)

XX Mouse Nkd protein gene sequence.

DE Mouse; gene; ds; gene therapy; Nkd; Wnt signaling pathway; cancer;

KM colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer;

XX murine.

OS Mus sp.

PN WO200292832-A2.

PD 21-NOV-2002.

PF 27-NOV-2001; 2001WO-US044092.

PR 27-NOV-2000; 2000US-0252884P.

PR 16-MAY-2001; 2001US-0291109P.

PR 01-OCT-2001; 2001US-0325571P.

XX (CHIR) CHIRON CORP.

XX Rohan M, Chan V, Yan D;

XX WPI; 2003-129303/12.

XX New human and non-human primate homologues of Nkd protein, and Nkd genes,

PT useful for treating cancer involving aberrant Wnt signaling, e.g. colon

PT cancer, head and neck cancer, ovarian cancer, or breast cancer.

XX Example 2; Fig 4; 99p; English.

XX The invention comprises the amino acid and coding sequence of the human

CC Nkd protein (hNkd) - a regulator of Wnt signaling pathways. The hNkd DNA

CC and protein sequences are useful for the treatment of cancer involving

CC aberrant Wnt signaling (e.g. colon cancer, head and neck cancer, ovarian

CC cancer and breast cancer). The present DNA sequence represents the mouse

CC Nkd gene

XX Sequence 1401 BP; 336 A; 430 C; 416 G; 219 T; 0 U; 0 Other;

SQ Alignment Scores:

Pred. No.: 6,49e-127 Length: 1401
 Score: 2171.50 Matches: 406
 Percent Similarity: 91.01% Conservative: 19
 Best Local Similarity: 86.94% Mismatches: 41
 Query Match: 86.89% Indels: 1
 Gaps: 1
 US-09-993-966-7 (1-470) x AAL51512 (1-1401)

QY 1 MetGlyLysLeuHisSerLysProAlaAlaValCysLeuArgArgLysSerProGlnGly 20
 |||||
 Db 1 ATGGGAAACCTTCACTCGAAGCCGCGCGGTGCAAGGCGAGGAGCCCGAAGGT 60
 |||||
 QY 21 AspSerPheAlaValSerAlaAlaTTPAlaArgLysGlyLleGlnGluTTPILeGlyArg 40
 |||||
 Db 61 GACAGCTTCTCTGTAAGCGCTGCTGGGCAAGAAAGCATCGAGAGTGGATCGGAGG 120
 |||||
 QY 41 GlnArgCysProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrLleGly 60
 |||||
 Db 121 CAGGCTGTCAAGCAGCGCTCTCAGAGCCCGTCAGCTGAGATTGGCAGGCACTGGT 180
 |||||
 QY 61 ArgSerThrArgGlyLeuValGlyAspValLeuArgAspThrLeuSerGlyGlnGlu 80
 |||||
 Db 181 CAGGCACTCGGAACTCGGTGACACTTCTAGAGAGGCTCTGGTGAAGAGAGAG 240
 |||||
 QY 81 AspAspPheArgLeuGlyValAlaLeuProProGlyLysThrAspGlyLeuGlySerGly 100
 |||||
 Db 241 GACGACTTCCCCCTTAAGAGTGGCCCTGCCCTGAGAAAGTCAACAGCTAGGTAGG 300
 |||||
 QY 101 AspGlyLysLysMetGlyArgValSerGlyProCysProGlySerLysLysGlnLeuLys 120
 |||||
 Db 301 GATGAGAAAGAAATGAGAGACTGAGGAGCACTGCGCAAGCCCTTCAAGAGAGCTCAG 360
 |||||
 QY 121 PheGlnGluLeuGlnCysAspValSerMetGlyLysSerArgGlnGluTTPThrPhe 140
 |||||
 Db 361 TTGGAAGGTCAAGTGTGTCCTGTGAGAGAGACGCGCAAGAGTGAAGCTTTC 420
 |||||
 QY 141 ThrLeuTyrAspPheAspAsnAsnGlyLysValThrArgLysLysAspLysThrSerLeu 160
 |||||
 Db 421 ACTTATATGACTTCGACCAACATGGCAAGTACCCGTAGACATACCAAGCTTCTG 480
 |||||
 QY 161 HisThrLleTyrGluValAlaAspSerSerValAsnHisSerProThrSerSerLysMet 180
 |||||
 Db 481 CATACCATCTATGAGGTGTTGACTCCTCTGTGAACCATTCCTCCCATCAAGCAACA 540
 |||||
 QY 181 LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysAspSerValLeuVal 200
 |||||
 Db 541 CTGGGGGTGAAGCTCACCGTCTCTGAGGAGGACCAAGTAAAGAGAGCGTCTTTTC 600
 |||||
 QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeu 220
 |||||
 Db 601 AACCATCCGATCTCGAGAGCAGAGCCCGCGAGCAGACACCAACCCGCTGAGAGACTG 660
 |||||
 QY 221 ArgSerTrpGlyLysLysGlnArgAlaProLeuArgPheGlnLysAspSerArgLeuGlu 240
 |||||
 Db 661 CGTGGCTGGAGAAAGAGCAGAGCCCACTCAGGTTCCAGGGGTGACAGCCACTGGAG 720
 |||||
 QY 241 GlnSerGlyCysTyrHisHisCysValAspGlyLysLleGluArgArgAsnHisLysLeu 260
 |||||
 Db 721 CAGGCAAGCTGACCAACCATTCGTGATGAAGAACTTGAAGGAGAAACCACTACTTA 780
 |||||
 QY 261 AspLeuAlaGlyLleGluAsnTyrThrSerGlnPheGlyProGlyLysSerProSerValAla 280
 |||||
 Db 781 GACCTGGGGGATAGAACTACACGTCATGTTGACCGGGATCCCTTCGTGGGCC 840
 |||||
 QY 281 GlnLysSerGlyLeuProProArgThrSerAsnProThrArgSerArgSerHisLysPro 300
 |||||
 Db 841 CAGAACTCAGAGCTGCCCTCGAATCTCAACCCCACTGCTCTGCTCCCAAGAGCA 900
 |||||
 QY 301 GlnAlaLleHisLysLeuProHisArgLysProGlnGlyValAspProAlaSerPheHisPhe 320
 |||||
 Db 901 GAAGCTGCCACATCCACACGAGGCGCCCAAGGTGAGACCAAGGCTCTTCACCTC 960
 |||||

QY 321 LeuapPThrProIleAlaValSerGluLeuGlnGlnIleuArgGlyThrGlnAsp 340
Db 961 CTGGACACCCCATTTGGCCAGGACATCCAGCAAGCGCTCCGGGCACTCAGAGAT 1020
QY 341 GlySerIleHisPheValArgSerProIleAlaGlnIleYsSerValGlyValGlyHis 360
Db 1021 GGGAGCAGAGACTTTGGAGGTCCCCCAAGGCCCAAGAGCAAGATGGGTATGGGCAC 1080
QY 361 ValIleAlaArgGlyAlaArgAnIleYsProIleuGlyProAlaIleProAlaValSerPro 380
Db 1081 GGGGCGCAGAGGTCCAGAGAACAGCCCTCCACTGGTACCCACCACTACTGCTCTCCCCC 1140
QY 381 SerAlaHisIleuAlaAlaSerProAlaLeuLeuProSerIleuAlaProIleuGlyHisIleYs 400
Db 1141 TCTGCCCATTTGGCCACACAGCCCAAGCCCTTCTCCCACTGGCAGCCCTGGGGCAGCAAG 1200
QY 401 LysHisIleHisIleArgAlaIleYsGluSerGlnGlnIleYsArgGlyIleuGlnAlaProIleu 420
Db 1201 AAACACAGCATGAGCCAGCCAGAGAGCCAGGACGCTCCGGGCTGCAAGGCCCTCTG 1260
QY 421 AlSerGlyGly--ProValIleuGlyArgGluHisIleuArgGluLeuProAlaLeuVal 439
Db 1261 GCTGACAGAGGCTCCACCGCTCATGGGGGGAGGAGGTGAGAGGCTGCTGCCGTGGTG 1320
QY 440 ValTyrGluSerGlnAlaGlyGlnProValGlnArgHisGluHisIleHisIleGlu 459
Db 1321 GTGTACAGAGCCAGGCTGGGCGGCGTCCAGAGACCAACACCATCACCACCA 1380
QY 460 HisHisIleHisIleTyrHisHis 466
Db 1381 CATCACCCATTTATCCACAC 1401

RESULT 6

AAL03214

AAL03214 standard; DNA; 2379 BP.

AAL03214;

21-NOV-2001 (first entry)

Human reproductive system related antigen DNA SEQ ID NO: 5902.

Human; reproductive system related antigen; reproductive system disorder;
cancer; gene therapy; ds.

Homo sapiens.

W0200155320-A2.

02-AUG-2001.

17-JAN-2001; 2001WO-US001339.

XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
XX 11-JUL-2000; 2000US-0216680P.
PR 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 22-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.

PR	08-NOV-2000;	2000US-0246533P.
PR	08-NOV-2000;	2000US-0246524P.
PR	08-NOV-2000;	2000US-0246525P.
PR	08-NOV-2000;	2000US-0246526P.
PR	08-NOV-2000;	2000US-0246527P.
PR	08-NOV-2000;	2000US-0246528P.
PR	08-NOV-2000;	2000US-0246532P.
PR	08-NOV-2000;	2000US-0246609P.
PR	08-NOV-2000;	2000US-0246610P.
PR	08-NOV-2000;	2000US-0246611P.
PR	08-NOV-2000;	2000US-0246613P.
PR	17-NOV-2000;	2000US-0249207P.
PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249209P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249215P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249217P.
PR	17-NOV-2000;	2000US-0249218P.
PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249264P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249297P.
PR	17-NOV-2000;	2000US-0249299P.
PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250160P.
PR	01-DEC-2000;	2000US-0250391P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
XX		
PA	(HUMA-)	HUMAN GENOME SCI INC.
XX		
PI	Rosen CA, Barash SC, Ruben SM,	
XX	WPI; 2001-465570/50.	
DR		
XX	Isolated nucleic acid molecule encoding a reproductive system antigen is	
PT	used in preventing, treating or ameliorating a medical condition.	
XX		
XX	Disclosure; SEQ ID NO 5902; 1297bp + Sequence Listing; English.	
FS		
XX		
CC	The present invention provides the protein and coding sequences of a	
CC	number of human reproductive system related antigens. These can be used	
CC	in the prevention and treatment of reproductive system disorders,	
CC	including cancer. The present sequence is a genomic sequence encoding a	
XX	protein of the invention	
XX		
QO	Sequence 2379 BP; 548 A; 689 C; 539 G; 603 T; 0 U; 0 Other;	

Alignment Scores:	
Pred. No.:	1,57e-57
Score:	1667.00
Percent Similarity:	58.08%
Best Local Similarity:	51.30%
Query Match:	42.70%
DB:	4
	Gaps: 16
US-09-993-966-7 (1-470)	x AAL03214 (1-2379)

QY	29	TPPLAAGdyrGlyIleGluGluThrIleGlyArgInaGyrProGlyGlyAlaSer	48
		:::	
Db	82	TGGGGATGAGGGCGAGGGCGTGGCGGAGCGGCGGGCGGTGGGGTGTCTAC	141
QY	49	GlyProArgInLeuArgLeuAlaGlyThrIleGlyArgSerThrArgGluLeuValGly	68
		:::	
Db	142	TGCTAACCCCAAGCTTCGTTAAGACACACTTATGGAACAGGTCAAG--CTTCTTGA	198
QY	69	Asp-ValLeuArgAspThrLeuSerGluGluGluLysAspAspPheArgLeuGluValAl	88
		:::	
Db	199	GAAGGTGACGTTAAAAATGTTTCAAAATTAAAG-----TAATCAATGACGAATATGCA	252
QY	88	AlaLeuProGluGlySerThrArgGlyLeuGlySerGlyLysArgLysLysMetGluArgVa	108
		:::	
Db	253	ACAAATATCAATAGTATCTAAAGCTTAAAGCCCAATGACAAACAGTTTGTGCCCC	312
QY	108	LSerGluProCys-----ProGly-----	114
Db	313	ATCCAGGCCCAACCCAGCCCTGTTCCCGGGGACACACTTGTACCTTCCTGGCTTTC	372
QY	115	SerLysLysGluLeuLysPheGluGluGluGlnCysAspValSerMetGluGluAspSe	134
Db	373	TTCACGAGAAACCTCGGTTTCCCTAAGTACAGCACTAGTGCTTTA-----	421
QY	134	rArgGlnGluThrPheThrLeuThrAspPheAspAsnGlyLysValThrArgL	154
Db	421	-----	421
QY	154	uAspLleThrSerLeuLeuHisThrLleTyGluValAlaAspSerSerValAsnHisSe	174
Db	422	-----TACGCTTGACATATGTTGACCTTGGCTCTCAATTTTCAC	459
QY	174	rProThrSer-----	177
Db	460	TCCCAACACTCCCTTCGCCCTTCTCTCAATGTTTAATTCTCTATGAGTACTTCT	519
QY	178	-----SerLysMetLeuArgValLysLeuThrValAlaProAsp-----	190
Db	520	GTATGTTTTTACATATATCTTAGAGT-----CTTCTTGCCCAAGTACTTAGTAGTT	576
QY	191	-----GlySerGlnSerLysAspSerValLeuVa	200
Db	577	TTCGGGGTCTCTGCTGCATCTGCTTCTCGGCTCCCTCTCGAG--AGATTTT--	631
QY	200	lAsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLe	220
Db	632	-----GAAAGCTTA-----CCAAAGAACAGTCTTCCCA-----CT	663
QY	220	uArgSerTrpGluLys-----LysGlnArgAlaProLeuArg-----	232
Db	664	TAGCAGATGGGAAAGTTCATTTACAGAGATTTTCTCTCAAGGACAGTAGGGGACGTACG	723
QY	233	-----PheGlnGlyAspSerArgLeuGlnGlnSerGlySerGlyHisLysCysValAspGl	251
Db	724	GGCATTTGGGGGTAGCTCTGCTTTTGGAGAACTGGGGGCG-----GGGGTATGTC	774
QY	251	uAsnIleGluArgAspAsnHisLysLeuAspLeuAlaGlyIleGluAsnTyThrSerGl	271
Db	775	TGGGGTATAGCGCAAGCCCAAGC-----ACAGTAAAGTGTCTATAATGTCTGTGA	828
QY	271	nPheGlyPro-GlySerProSerValAlaGlnLysSerGluLeuProProArgThrSerA	291
Db	829	ATTGGTTCTTAGGCTCCCTTCCTCGGCGCCAGAACTGAGAACTGCCCCCGCAGCTTCCA	888
QY	291	snProThrArgSerArgSerHisGluProGluAlaIleHisLleProHisArgLysProG	311
Db	889	ATCCCACTCGATCTGCTCCATGAGCGGAGGCATCCATCCACACCGAAGAGCCCC	948
QY	311	lGlnIValAspProAlaSerPheHisPheLeuAspThrProIleAlaLysValSerGluL	331
Db	949	AAAGGGTGAACCCGGGCTCTTCCATCTCTTGAACACCCCATCGGCAAGGTCTAGACG	100
QY	331	eGlnGlnArgLeuArgGlyThrArgInaArgLysSerLysHisPheValArgSerProLysA	351

Db 1009 TCCAGCAACGGCTCCGGGGACCCAGGACGGAGCAACCTTTGTAGGTCCCCAAAG 1068
Oy 351 IagInglYlySerValGIyValGIyHlValAlaArgAlaArgAsnLysProProl 371
Db 1069 CCGAGGGCAAGAGTGTGGTGTGGGCCACGTGGCCAGAGGGGCAAGAAACAAGCCCCCTC 1128
Oy 371 euGlyProAlaIleProAlaValSerProSerAlaHlSerProAlaLeul 391
Db 1129 TGGGACCCCGCATCCCTGGGTGTCTCCCTCCGCGCACCTGGCCGAGCCGCTCC 1188
Oy 391 euProSerLeuAlaProLeuGIyHlSerLysHlSerHlSerAlaLysGIuSerGIg 411
Db 1189 TCCCTCTCCCTAGCCCTCCCTGGGCAAGAGCAAGACCAAGGAGAGCCAGC 1248
Oy 411 InGIyCysArgGIyLeuGIuAlaProLeuAlaSerGIyGIyProValLeuGIyArgGIuH 431
Db 1249 AGGGCTGCGGGGCGCTGCAGGCACTGGCTCAGGTGGCCCTGTCTGGGGCGGAGC 1308
Oy 431 IAlaLeuArgGIuLeuProAlaLeuValIyValIyGIuSerGIuAlaGIyGIuProValGIuAla 451
Db 1309 ACCTGCGGAGCTGCTCCGCTGTGTGTATGAGCCAGGCGGCGCAGCCGCTCCAGA 1368
Oy 451 rGHIaGIuHlSerHlSerHlSerHlSerHlSerHlSerHlSerHlSerHlSerHlSerHlSerHl 470
Db 1369 GACATGAGCACCAACCAACATGACATCAACCAATACCACTTACCAAGACA 1427

RESULT 7
AAL03216
ID AAL03216 standard; DNA; 2379 BP.
XX AAL03216;
AC
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 5904.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN MO200155320-A2.
XX
PD 02-AUG-2001.
XX
XX
PF 17-JAN-2001; 2001MO-US001339.
XX
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205151P.
PR 07-JUN-2000; 2000US-0209467P.
PR 30-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 26-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 14-AUG-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234988P.
PR 26-SEP-2000; 2000US-0234989P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236680P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.

PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249247P.
PR 17-NOV-2000; 2000US-0249256P.
PR 17-NOV-2000; 2000US-0249257P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250319P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 06-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI, 2001-465570/50.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.
XX
XX
PS Disclosure; SEQ ID NO 5904; 1297PP + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention
XX
SQ Sequence 2379 BP; 548 A; 689 C; 539 G; 603 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.57e-57 Length: 2379
Score: 1067.00 Matches: 257
Percent Similarity: 58.08% Conservative: 34
Best Local Similarity: 51.30% Mismatches: 100
Query Match: 42.70% Indels: 110
DB: 4 Gaps: 16

US-09-993-966-7 (1-470) x AAL03216 (1-2379)
QY 29 TTPALARGLYSGLYLEGLUGLUTRPILLEGLYARGGLNARGCYSPROGLYGLYVALSER 48
DB 82 TGGGCGATGAGGGGAGGGCGCTGAGCGGGCCAGGGGCGCGTGGGGGTGTTTAC 141
QY 49 GLYPROARGLNLEUARGLEUVALGLYTHRILLEGLYARGSERTHRARGLLEUVALGLY 68

DB 142 TGCTACCCAGAGCTTGTAAGACACTATTATGGACAGGTCAAGA---CTTCTTGA 198
QY 69 Asp-ValLEUARGAspThrLEUSerLUGLUGLUGLUAAspPheARGLEUVAL 88
DB 199 GAAGGTGAGCTTAAATGGTTTCMAAATTAAG-----TAATGATGCACAAATCGA 252
QY 88 aleuPProGLULYThrAspGLYLeuGLYSerGLYAspGLULYLeuMetGLUARGVA 108
DB 253 ACAAATCCAAATGTAATCTTAAGGCTTAGACCCCAATGACAAACACAGTTTGTGCCCC 312
QY 108 lSerGLUProCYs-----ProGLY----- 114
DB 313 ATCCAGCCCAACCCAGCCCTGTCCGGGGCAACCACTTGACCTCTCCGTGCTTTTC 372
QY 115 -SerLYSGLINLEUArgPheGLULeuGLINCYsAspVALSerMetGLULAspS 134
DB 373 TTCACCAAGAACCTGGTTCCCTAAGTACAGACACTACTGTGTCTTTA----- 421
QY 134 rArgGLNGLUTrPThrPheThrLeuTYrAspPheAsnAsnGLYValThrArgL 154
DB 421 ----- 421
QY 154 uAspILEThrSerLeuLeuHIePhrILETYrGLUVALAspSerSerVALenHISe 174
DB 422 -----TCAGCCTGACATATGTTGACCTTGCTCAATTTAC 459
QY 174 rProThrSer----- 177
DB 460 TCCCAACAATCCCTTCCCTCTCTCTCATATGTTAATCTCTATGACTTCTT 519
QY 178 -----SerLYMetLeuArgVALYsLeuThrVALAspAsp----- 190
DB 520 GTATGTTTTTAAAGAAATATACTTAGATT---CTTCTGCCCAAGGACCTTAGATATT 576
QY 191 -----GlySerGLINSerLYsArgSerVALeuVA 200
DB 577 TTCGGGGTCTCTGTCGACATCTGCTTTCGAGCTCCCTCGAG---AGTATTTT 631
QY 200 lAsnGLNALAspLeuGLINSerVALArgProVALAspGLUThrLYPProThrGLUAspLe 220
DB 632 -----GAAGACCTA-----CCAAAGAACCACTTCTTCCA-----CT 663
QY 220 uArgSerTrpGLULys-----LYsGLNArgVALProLeuArg----- 232
DB 664 TAGCAGATGGAAAGTTCATTCACAGAGTTTCTTCGCAAGCAGAGGGGACATCAG 723
QY 233 ----PheGLNGLYAspSerArgLeuGLINSerGLYCYeTYrHIeHISCYsVALAspGL 251
DB 724 GGCATTGGGGGTAGCTCTGTTTGGAGAACTGGGGGCG-----GGGGTGAATGTC 774
QY 251 uAsnILLEGLUArgArgAsnHISrTYrLeuAspLeuVALAspLYLeuAsnTYrHISerGL 271
DB 775 TGGGGTATAGCGGAAGCCCAAGCAGC-----ACAGTAGGGTCAATTAATGCTGTTGA 828
QY 271 nPheGLYPro-GlySerProSerVALAspGLINLYsSerGLULeuProProArgThrSerA 291
DB 829 ATTGTTCTTAAGCTCTCCCTTCCGTGGCCAGAAAGTCAAGATGCCCGCCGACCTCCA 888
QY 291 snProThrArgSerArgSerHISGLUProGLUALIeHISILEProHISArgLYsProG 311
DB 889 ATCCCAATGATGATCTCGTCCATGAGCGGAAACCATCCACATCCACAGCAAGAGCCC 948
QY 311 lNGLYVALAspProAlAserPheHISrPheLeuAspThrProILEALYValSerGLUL 331
DB 949 AAGGCGTGAACCCGGCTCTTCCACTCTTACACCCCAATCGCCAAAGTCTCAAGC 1008
QY 331 eUGlNGlNArgLeuArgGLYThrGLINAspGLYSerLYsHISpHeVALArgSerProLYsA 351
DB 1009 TCCAGCAACGAGCTCCGGGGCAACCCAGACGGGAGCAAGCACTTGTGATGATCCCCAAG 1068
QY 351 lAslNGLYLSerSerVALYValGLYHISVALAlAspArgLYAlAspAsnLYsPProL 371

XX WPI: 2001-639362/73.
 DR P-PSDB; ABG14565.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 PS Claim 1; SEQ ID NO 14556; 103bp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridization probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptides and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1743 BP; 367 A; 562 C; 511 G; 303 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1,32e-43 Length: 1743
 Score: 842.00 Matches: 196
 Percent Similarity: 43.71% Conservative: 9
 Best Local Similarity: 41.79% Mismatches: 13
 Query Match: 33.69% Indels: 251
 DB: Gaps: 5

US-09-993-966-7 (1-470) x AAS76752 (1-1743)

QY 9 AAlAlAValCyblySArGArGluSerProGlu----- 19
 DB 12 GCCGCGGTGTGCAAGCCAGAGAGAGCCGGAAGTAAAGGCGCGGCGCAGACCTC 71
 QY 20 ----- -GlyAASPSe 22
 DB 72 GGGGATGAGCGGGGGGACACCGCGCGCGGCAAGCCGCCCGCCAGGTGACAG 131
 QY 22 rPhAlAValSerAlaAlaTrpAlaArgLygLyIleGluGluTrpIleGlyArgGlnAr 42
 DB 132 CTTCGCGGTGAGCGCTCGTGGCTCGGAAGGCGATCGAGAGTGGATCGGAGACAGCG 191
 QY 42 gCySProGlyGlyValSerGlyProArgGluLeuArgLeuAlaGlyThrIleGlyArgSe 62
 DB 192 CTGCGCGGCGGTCTCTCGGACCCGACAGCTGCGGTGGCGGCAACATAGGCCGAG 251
 QY 62 rThArgGluLeuValGlyAaSPValLeuArgSPThrLeuSerGluGluGluAaSPa 82
 DB 252 CACCCCG----- 258
 QY 82 pPhArgLeuGluValAlaLeuProProGluLySThrAaSPGlyLeuGlySerGlyAaSPG 102
 DB 258 ----- 258
 QY 102 uLybLyMetGluArgValSerGluProCySProGlySerLybGluLeuLySPheG 122
 DB 258 ----- 258

QY 122 uGluLeuGlnCyAaSPValSerMetGluGluAaSPerArgGlnGluTrpThrPheThrLe 142
 DB 259 -----ARGTCTGTGAGAGTCCCG----- 276
 QY 142 uTyAaSPheAaSPaAaSPGlyValThrArgGluAaSPLeThrSerLeuLeuHsrh 162
 DB 276 ----- 276
 QY 162 rIleTyGluValValAaSPerSerValAaSPHrSerProThrSerSerLySMetLeuAr 182
 DB 276 ----- 276
 QY 182 gValLybLeuThrValAlaAaSPArgLySerGlnSerLyAaSPerValLeuValaSPG 202
 DB 277 -----AACTGTGACGGGCTTCAGGACGTGCGGTGAACT 314
 QY 202 nAlaAaSPLeuGlnSerAlaArgProArgAlaGluThrLybProThrGluAaSPArgSe 222
 DB 315 CGGCAAGCCCAAGGCGC----- 333
 QY 222 rTrpGluLybGlnArgAlaProLeuArgPheGlnGlyAaSPerArgLeuGlnSe 242
 DB 333 ----- 333
 QY 242 rGlyCySTyHrHsrhCyAaSPValAaSPGluAaSPLeGluArgAaSPHrTyGlyAaSPLe 262
 DB 333 ----- 333
 QY 262 uAlaGlyIleGluAaSPTrhSerGlnPheGlyProGlySerProSerValAlaGlnLy 282
 DB 334 -----TCCCTTCCTGTGGCCCAAGAA 353
 QY 282 sSerGluLeuProProArgTrhSerAaSPProThrArgSerArgSerHsrhGluProGluAl 302
 DB 354 GTCAAGACTGCCCCCGCCGACCTCCAAATCCCATCTCATCTCCCATAGCCGGAAGC 413
 QY 302 aAlenHsrhProHsrhArgAaSProGlnGlyValAaSPProAlaSerPheHsrhAaSP 322
 DB 414 CATCCACATCCCAACAGGAAAGCCCAAGGCGTGAACCGGCTCTCCACTCCCTTGA 473
 QY 322 rThrProIleAlaLybValSerGluLeuGlnGlnArgLeuArgGlyThrGlnAaSPGlySe 342
 DB 474 CACCCCAATGCCAAGGTCTCAAGCTCCAGCAACGCTCCGGGGCAACCAAGACGGGAG 533
 QY 342 rLybHsrhAaSPValArgSerProLybAlaGlnGlyLySerValGlyValGlyHsrhAaSP 362
 DB 534 CAAGCACTTGTGAAGCCCCCAAGGCCCAAGGCCAAGAGTGTG----- 576
 QY 362 aArgGlyAlaArgAaSPHrProLeuGlyProAlaIleProAlaValaSPerProSerAl 382
 DB 576 ----- 576
 QY 382 aHsrhLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlnLyHsrhLybHsrh 402
 DB 577 -----GCCCTTCGGGCAAGAAAGCA 599
 QY 402 sLybHsrhAaSPAlaLybGluSerGlnGlnGlyCyAaSPGlyLeuGlnAlaProLeuAlaSe 422
 DB 600 CAAGCAACGAGCCAGAGAGGCAAGGAGGCTGCGGGGCTCGGACACCATGCGCTC 659
 QY 422 rGlyGlyProValLeuGlyArgGluHsrhAaSPArgGluLeuProAlaLeuValValTyG 442
 DB 660 AGGTGGCCCTGTCTGGGAGGAGGACCTGCGGAGCTCCGCTGTGTGTGTATGA 719
 QY 442 uSerGlnAlaGlyGlnProValGln 450
 DB 720 GAGCCAGGCCGGGACGCCGCGCAG 744

RESULT 10
 AAS92081
 ID AAS92081 standard, cDNA; 1743 BP.
 XX
 AC AAS92081;

QY 382 aHisLeuAlaIaSerProAlaLeuProSerLeuAlaProLeuGlyHisLeuHis 402
 Db 577 -----GCCCTCTGGGCAAGCA 599
 QY 402 sLYSHiARgAlaLYSGiuserGInGlyCYaRGiLYeUGiAlaProLeuAlaSe 422
 Db 600 CAAGCACCAGGCCAAGGAGGACGACGGCTGCGGGGCTTGACAGGACCACTGGCTTC 659
 QY 422 rGLYGLYProValLeuGLYARGLYUHiSLeUARGiLUeUProAlaLeuValIYRGi 442
 Db 660 AGGAGGCTCTGCTCTGGGGGCGGAGCACTGCGGAGCTGCGCTTGAGTGTATGA 719
 QY 442 uSERGLiAlAGiLYGInProValGIn 450
 Db 720 GAGCCAGGCGCGGACGCGGCCAG 744
 RESULT 11
 ID ADC31966
 ID ADC31966 standard; cDNA; 1743 BP.
 AC ADC31966;
 DT 18-DEC-2003 (first entry)
 XX
 DE Human novel cDNA contig sequence, SEQ ID NO:2048.
 XX
 XX Human, diagnostic; drug screening; forensics; gene mapping;
 KM biodiversity assessment; Parkinson's disease; Alzheimer's disease;
 KM neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
 KM ulcers; osteoporosis; autoimmune disease; cancer;
 KM molecular weight marker; food supplement; antiparkinsonian; nootropic;
 KM neuroprotective; antihaemic; anticoagulant; thrombolytic; vulnery;
 KM antitumor; osteopathic; immunosuppressive; antiinflammatory; cyrostatic;
 KM gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO2003029271-A2.
 XX
 PD 10-APR-2003.
 XX
 PF 24-SEP-2002; 2002WO-US030474.
 XX
 PR 24-SEP-2001; 2001US-0324631P.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
 PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
 PI Haley-Vicente D, Drmanac RT;
 XX
 XX WPI; 2003-371981/35.
 DR P-PSDB; ADC32733.
 XX
 XX New polynucleotide and polypeptide useful for diagnosing, preventing or
 PT treating conditions such as neurodegenerative diseases, anemias, platelet
 PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
 PT cancer.
 PS
 PS Example 2; SEQ ID NO 2048; 1185bp; English.
 XX
 XX The invention relates to 971 novel human cDNA sequences (ADC29919-
 CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
 CC invention also relates to nucleic acid sequences over 99% identical with
 CC the novel human cDNAs. The invention additionally encompasses expression
 CC vectors and host cells comprising a nucleic acid of the invention; the
 CC recombinant production of a polypeptide of the invention; an antibody
 CC against a polypeptide of the invention; a method of detecting
 CC polynucleotides or polypeptides of the invention; and methods of
 CC identifying a compound which binds to a polypeptide of the invention. The
 CC invention further discloses methods of preventing, treating or
 CC ameliorating a medical condition; kits comprising polynucleotide probes

CC and/or monoclonal antibodies for carrying out the methods of the
 CC invention; methods for the identification of compounds that modulate the
 CC expression or activity of the polynucleotide and/or polypeptide; and 767
 CC contig sequences corresponding to the cDNA sequences of the invention
 CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
 CC -ADC33394). The nucleic acids and polypeptides of the invention are
 CC useful in diagnostics, drug screening, forensics, gene mapping, in the
 CC identification of mutations responsible for genetic disorders or other
 CC traits, for assessing biodiversity, and in producing many other types of
 CC data and products dependent on DNA and amino acid sequences. They are
 CC also used for treating diseases such as Parkinson's disease, Alzheimer's
 CC disease and other neurodegenerative diseases, anaemia, platelet
 CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
 CC cancer. The nucleic acids may also be used as hybridization probes or
 CC primers, and in the recombinant production of a protein. The polypeptides
 CC are also useful in generating antibodies, as molecular weight markers,
 CC and as food supplements. The present sequence represents a human contig
 CC sequence used in an example of the invention. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1743 BP; 367 A; 562 C; 511 G; 303 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1,32e-43	Length:	1743
Score:	842.00	Matches:	196
Percent Similarity:	43.71%	Conservative:	9
Best Local Similarity:	41.79%	Mismatches:	13
Query Match:	33.69%	Indels:	251
DB:	10	Gaps:	5

US-09-993-966-7 (1-470) x ADC31966 (1-1743)

QY 9 ALaAlaValCYsLYaRgATgUSeRProGlu----- 19
 Db 12 GCCGCCGTGTGCAAGCCGAGGAGAGCCCGAAAGTAGGGCGCGCGGCGAGACCTC 71
 QY 20 -----GlyAsPSe 22
 Db 72 GGGGATGACGCGGGGGACACCGCGCGCGGCAAGCGCCAGCCCGCAGGTGACAG 131
 QY 22 rPHeAlaValSeRAlaAlaTrPaIARGLYSGiLYIeGLUGiUTrPIeGLYARGiNAR 42
 Db 132 CTTCGCCGTGAGCGCTGCTCGGGCTCGGAAGGCAATGAGATCGGAGACAGCG 191
 QY 42 gCYSPROGLiYLYaLSerGLYProARGLiNLeUARGiLUeAlAGLYThrIleGLYARSe 62
 Db 192 CTGCCCCGGGGGIGTCTCGGGACCCCGACAGCTGCGGTGGCGGGCACATAGCCGAG 251
 QY 62 rThARgGLiUeUValGLYAsPValLeUARGAsPThLeUSeRGLUGiUAsPAs 82
 Db 252 CACCCGG----- 258
 QY 82 pHeARGLeUGiUAlaLeUProProGLiUYSThRAsPGLYLeUGLYSeRGLYAsPGL 102
 Db 258 ----- 258
 QY 102 uLYsLYSeMeGLUARGValSeRGLUPROCYSPROGLYSeRLYSeGLiNLeUysPHeGL 122
 Db 258 ----- 258
 QY 122 uGLiUeUGiCYAsPAsPValSeRMeTGLUGiUAsPSeRARGiNGLiUTrPhRThLe 142
 Db 259 ---ATGTCTGTGAGCTCCG----- 276
 QY 142 uTYRAsPHeAsPAsnAGiLYsValThRARGiUAsPILeThSeRLeUeUHiSTH 162
 Db 276 ----- 276
 QY 162 rILETYRGLiUAlaValAsPSeSeRValAsnHiSeRProThSeRSeRlySeMeTLeUAr 182
 Db 276 ----- 276

QY 182 gValylsLeuThrValAlaProAspGlySerGlnSerIleArgSerValIleuValaInGI 202
Db 277 -----AACTCTGACCGGGCTTTCAAGCACTGCGCGAAGT 314
QY 202 nAlaAspLeuGlnSerAlaArgProArgAlaGluThrIleProThrGluAspLeuArgSe 222
Db 315 CGGCGAAGCCCAAGGCAAGC----- 333
QY 222 rTPrGluLysLeuGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlnIse 242
Db 333 ----- 333
QY 242 rGlyCysTyrHisIleCysValaAspGluAsnIleGluArgArgAsnIleTyrLeuAspLe 262
Db 333 ----- 333
QY 262 uAlaGlyIleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValaIaGlnIly 282
Db 334 -----TCCCTTCCTGCGCCCAAG 353
QY 282 sSerGluLeuProProArgThrSerAsnProThrArgSerArgSerHisGluProGluAl 302
Db 354 GTCAGAACTGCCCCCGCACCTCCATCCCATCTCGCTCCCATGAGCGCGAAGC 413
QY 302 aIleHisIleProHisArgIleProGlnGlyValaAspProAlaSerPheHisPheLeuAs 322
Db 414 CATTCACATCCCAACCGAAAGCCCCAAGCGCTGAGCCCGCTTCCTCCATCTTCCTTGA 473
QY 322 pThrProIleAlaIleValaSerGluLeuGlnIleArgLeuArgGlyThrGlnAspGlySe 342
Db 474 CACCCCAATGCCAAGGTCTCAAGCTCCAGCAAGCGCTCGGGGCAACCAAGAGCGGAG 533
QY 342 rIlyHisIlePheValaArgSerProLysAlaGlnIlyLysSerValaGlyValaGlnHisValaI 362
Db 534 CAAGCACTTGTGTAGAGCCCCCAAGGCCCAAGGCAAAAGTGT----- 576
QY 362 aArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerProSerAl 382
Db 576 ----- 576
QY 382 sHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlnIlyHisLysIlyHis 402
Db 577 -----GCCCTCCCGGCAAGAAAGCA 599
QY 402 sIlyHisIleArgAlaIleGluSerGlnGlnIlyCysArgIlyLeuGlnAlaProLeuAlaSe 422
Db 600 CAAGCAACGAGCCAAAGAGAGCCAGGAGCTGCGGAGCTGCAAGCACACAGCTGCTC 659
QY 422 rGlyGlyProValaLeuGlyValArgGluHisIleLeuArgGlyLeuLeuProAlaLeuValIlyrGI 442
Db 660 AGGTGACCTGTCTCTGGGGGAGAGCACTGCGGAGCTGCGGCTGTGTGTGTATGA 719
QY 442 uSerGlnAlaGlyGlnProValaGln 450
Db 720 GAGCCAGGCCGAGGAGCCGAGCCAG 744

RESULT 12
AAS27117
ID AAS27117 standard; cDNA, 1966 BP.
XX AAS27117;
XX
DE 07-NOV-2001 (first entry)
XX
XX cDNA encoding novel signal transduction pathway protein, Seq ID 152.
XX
XX Neuroprotective; cytoskeletal; dermatological; immunosuppressive; tumour;
XX antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
XX immune system disorder; rheumatoid arthritis; inflammatory condition;
XX organ transplant rejection; infection; hepatitis C; blood disorder;
XX sickle cell anemia; hyperproliferative disorder; Gaucher's disease;
XX neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;

KM chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KM cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KM reproductive system; gastrointestinal; liver disorder; AIDS; ss;
KM acquired immune deficiency syndrome.
XX
XX Homo sapiens.
XX
XX W0200154733-A1.
PD
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001MO-US001312.
PF
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226682P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.

PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236337P.
PR 29-SEP-2000; 2000US-0236357P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 08-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249267P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.

PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254079P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465460/50.
DR P-PSDB; AAU17200.
XX
XX Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders.
XX
PS Claim 1, SEQ ID NO 152; 880pp; English.
XX
XX The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections, and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative disorders
CC (e.g. Gaucher's disease and cancer), neurodegenerative disorders (e.g.
CC Alzheimer's disease, Parkinson's disease), chromosomal abnormalities
CC (Down syndrome), ischaemic injury (e.g. stroke), renal disorders (e.g.
CC glomerulonephritis), cardiovascular disorders (e.g. arrhythmia),
CC respiratory disorders, dermatological disorders, in wound healing,
CC epithelial cell proliferation, endocrine disorders (e.g. Addison's
CC disease), reproductive system disorders, gastrointestinal disorder
CC (inflammatory disorders), liver disorders (cirrhosis), as stimulators of
CC B-cell responsiveness to pathogens, activators of T-cells, to induce
CC higher affinity antibodies, and as a means to induce tumour proliferation
CC in pathologies e.g. acquired immune deficiency syndrome (AIDS). AAS26976-
CC AAS27850 represent novel signal transduction pathway protein coding
CC sequences and PCR primers of the invention
XX
Alignment Scores:
Pred. No.: 6.91e-43 Length: 1966
Score: 831.50 Matches: 218
Percent Similarity: 51.74% Conservative: 49
Best Local Similarity: 42.25% Mismatches: 142
Query Match: 33.27% Indels: 107
DB: Gaps: 19
US-09-993-966-7 (1-470) x AAS27117 (1-1966)
QY 1 MetGlyLysLeuHisSerLysProAlaValAlaCys---LysArgArgGluSerProGlu 19
DB 252 ATGGGGAACCTGCAATCGAAGCAAGCCGCCGCCGCGCAAGAGGAGAGAGAGCCGAA 311
QY 20 -----GlyAspSerPheAlaValSeraLa---Ala 28
DB 312 GGTTCCTCCGGCGCTCCGCCCGGACCGCAGAGGACAGCTTCGTGGCGTCCGCTACGCG 371
QY 29 TrpAlaArgLysGlyIleGluGluTrpIleGlyArg-----GlnArgCysPro 44
DB 372 AGCGGCGCGCAAGCGCGGAGGAAGCGGAGCGCGCGCGGAGCAAGAGAGCTCCCC 431
QY 45 GlyGly-----ValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGlyArgSer 62
DB 432 AATGGGAGCCCAAGAGAGGCGGCTTCGCG----- 461
QY 63 ThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGluAspAsp 82
DB 462 -----GAGGACCAAG 470
QY 83 PheArgLeuGluValAlaLeuProProGluLysThrAspGly----- 96


```
Db 471 TGTCCCTAGAGGTGGACCTCCCGCTGAGAAAGCTGAGGGCCCGAGCACCAGGACCA 530
Qy 97 ---LeuGlySerGlyAap---GluLysLysMetGluArgValSerGluProCysArgGly 114
Db 531 CTCTCTCAGCCAGATGACGAGAGAGAGGACAAACCGGAGGGCCCGGAGACCGGGAC 590
Qy 115 SerLysLysGlnLysLysPheGlnGlnLysGlnLysPheValSerMetGlnLysSer 134
Db 591 GGGAGGGC---CTCAACATTGACGACCTCCAGTGCATGCTGAGAGAGAGAGAC 647
Qy 135 ArgGlnGlnTrpThrPheThrLeuTyrAspPheAspPheAsnGlnLysValThrArgGlu 154
Db 648 CGCCAGGAGTGAAGCTTCAAGCTTATGACTTTGACAACTCCGGGAAGTCAACAGGAG 707
Qy 155 AspLleThrSerLeuLeuLysIleThrIleTyrGluValValAspSerSerValAsnHisSer 174
Db 708 GACATGTCACAGCTCATGACACACCATCTATGAGGTCTGATGCTGCTGCTCAACACTCC 767
Qy 175 ProThrSerSerLysMetLeuArgValLysLeuThrValAlaProAspGlySerGlnSer 194
Db 768 TCGGGCAGCAGCAAGACCTTCGTGTAAAGTAAACCGTCAGCTTCCAGCTCCAGCAAG 827
Qy 195 LysArgSerValLeuValAsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThr 214
Db 828 AGGAAGAGGGTCTCTCTGCTGGCCAGAGCCGGAGCCCAACCTTTCAGATGAGAGGT 887
Qy 215 LysProThrGluAspLeuArgSerTrpGluLysGlnArgAlaProLeuArgPheGln 234
Db 888 GAACCTGACAGAGGAGCCAGGGGTGCTGACAGAGGTTGCTGTCACAGTCAGAGAGCCC 947
Qy 235 GlyAspSerArgLeuGlnLysGlnSerGlyCysTyrHisLysCysValAspGluAsnLysGlu 254
Db 948 AGTACGACCCCCAGCCCTGCTGGAGCGGGGGCCCTGACCTGCTGACAGAAACAGGAG 1007
Qy 255 ArgArgAsnHisTyrLeuAspLeuAlaGlyIleGluAsnTyrThrSerGlnPheGlyPro 274
Db 1008 CGCAGAAACACACTACCTGGAACCTGCGCGGATTTAGAACTACACTCCAGATTGCGCCCT 1067
Qy 275 GlySerProSerValAlaGlnLysSerGluLeuProArgThrSerAsn---ProThr 293
Db 1068 GGGTCCCTCTCTGCAAGCAAGACAGAGGCCCGGAGGCTCGACCTCCAGAGCC 1127
Qy 294 ArgSerArgSerHisGluProGluAlaIleHisIleProHisArgGlyProGln----- 311
Db 1128 CGGTCCCGCTCCAGAGCAGCATACACATGCCCTACACACCGCAGCTCAGCTGCTG 1187
Qy 312 -----GlyValAspProAlaSerPheHisPheLeuAspThrProIleAlaLysValSer 329
Db 1188 GTGGAACAGCTCGTGCAGCTGCG-----GAGCTGCTGCTCGCGGGCCCTG 1232
Qy 330 GlnLeuGlnGlnArgLeuArgGlyThrGlnAspGlySerLysIlePheValArgSerPro 349
Db 1233 GACACGAGCGCCCGCGGAGGGGCGGAG-----AAGCAGTTCTTCAGTCCGCC 1283
Qy 350 LysAlaGlnLysLysSerValGlyValGlnHisValAlaArgGlyAlaArgAsnLysPro 369
Db 1284 AAGGGCTCCGGGAAGCGG-----1301
Qy 370 ProLeuGlyProAlaIleProAlaValSerProSer-----AlaHisLeuAlaAlaSer 387
Db 1302 -----CTTGGGGTGCACAGCAGCAAGTCCGGAAAGCTTCACTACTACTG 1352
Qy 388 ProAlaLeuLeuProSerLeuAla-----395
Db 1353 CCGGCGCTCTCGCGCCAGGCGCTCAGAGAGCGCCACACCTCCGAGCGCCACCG 1412
Qy 396 ---ProLeuGlyLysLysLysHisLysHisArgAlaLysLysGlnGlnGlnGlyCysArg 414
Db 1413 CCAACCTACCGCAGCAAGCGGTACCGCAAAAGGCGAGGAG-----GGCCTCTG 1463
Qy 415 GlyLeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlnLysArgLysHis-----Leu 432
Db 1464 CCACTCAAGGCCCAACAGCTCAGCTCAGCTGCAAGCTG-----GAGCTCAGAGGTG 1514
```

```
Qy 433 ArgGluLeuProAlaLeuValTyrGluSerGlnAlaGlyGlnProValGlnArgHis 452
Db 1515 CGGACCTGCGCGCCACGACGAGAGAGGGCTACCGGTGCGCATGATCCAGCGGAC 1574
Qy 453 GlnHisHisHisHisHisGlnHisHisHisHisHisHisHisHisHisHisHisHis 468
Db 1575 GAGCACCACACACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1622

RESULT 13
ADB93295
ID ADB93295 standard; cDNA, 1966 BP.
AC ADB93295;
DC
DT 04-DEC-2003 (first entry)
XX
DE Human cDNA encoding a novel protein #142.
XX
KW 89; gene; human; autoimmune disease; Parkinson's disease; silleosis;
KW gastrointestinal disease; atherosclerosis; haemophilia; thrombocytopenia;
KW immunosuppressive agent; adjuvant; enhance immune response;
KW higher affinity antibody induction;
KW increased serum immunoglobulin concentration.
XX
OS Homo sapiens.
XX
PN US202168711-A1.
XX
PD 14-NOV-2002.
XX
17-JAN-2001, 2001US-00764868.
XX
31-JAN-2000, 2000US-0179065P.
PR 04-FEB-2000, 2000US-0180628P.
PR 26-JUN-2000, 2000US-0214886P.
PR 07-JUL-2000, 2000US-0216647P.
PR 07-JUL-2000, 2000US-0216880P.
PR 11-JUL-2000, 2000US-0217487P.
PR 11-JUL-2000, 2000US-0217496P.
PR 14-JUL-2000, 2000US-0218290P.
PR 26-JUL-2000, 2000US-0220963P.
PR 26-JUL-2000, 2000US-0220964P.
PR 14-AUG-2000, 2000US-0224518P.
PR 14-AUG-2000, 2000US-0224519P.
PR 14-AUG-2000, 2000US-0225267P.
PR 14-AUG-2000, 2000US-0225268P.
PR 14-AUG-2000, 2000US-0225270P.
PR 14-AUG-2000, 2000US-0225447P.
PR 14-AUG-2000, 2000US-0225757P.
PR 14-AUG-2000, 2000US-0225758P.
PR 22-AUG-2000, 2000US-0226688P.
PR 30-AUG-2000, 2000US-0228924P.
PR 01-SEP-2000, 2000US-0229287P.
PR 01-SEP-2000, 2000US-0229343P.
PR 01-SEP-2000, 2000US-0229345P.
PR 01-SEP-2000, 2000US-0229345P.
PR 05-SEP-2000, 2000US-0229509P.
PR 05-SEP-2000, 2000US-0229513P.
PR 08-SEP-2000, 2000US-0231413P.
PR 21-SEP-2000, 2000US-0234223P.
PR 21-SEP-2000, 2000US-0234274P.
PR 25-SEP-2000, 2000US-0234997P.
PR 27-SEP-2000, 2000US-0235834P.
PR 29-SEP-2000, 2000US-0236327P.
PR 29-SEP-2000, 2000US-0236367P.
PR 29-SEP-2000, 2000US-0236368P.
PR 29-SEP-2000, 2000US-0236369P.
PR 29-SEP-2000, 2000US-0236370P.
PR 02-OCT-2000, 2000US-0236802P.
PR 02-OCT-2000, 2000US-0237037P.
PR 02-OCT-2000, 2000US-0237038P.
PR 02-OCT-2000, 2000US-0237039P.
```

PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 20-OCT-2000; 2000US-0240950P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251858P.
 PR 08-DEC-2000; 2000US-0251869P.
 XX
 PA (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 PI Rosen CA, Ruben SM, Barash SC,
 XX
 XX WPI; 2003-719985/68.
 DR P-PSDB; ADB93908.
 DR
 XX
 PT New isolated polypeptide useful for diagnosing and treating
 PT immunosuppressive conditions such as autoimmune disease and Parkinson's
 PT disease.
 PS
 PS Claim 3; SEQ ID NO 152; 345pp; English.
 XX
 CC The invention relates to an isolated polypeptide. The polypeptide is
 CC useful for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition in a subject, by determining the presence or
 CC amount of expression of the polypeptide in a biological sample and
 CC diagnosing a pathological condition or a susceptibility to a pathological
 CC condition based on the presence or amount of expression of the
 CC polypeptide. The polypeptide is also useful for identifying a binding
 CC partner to the polypeptide, which involves contacting the polypeptide
 CC with a binding partner and determining whether the binding partner
 CC effects an activity of the polypeptide. The polypeptide or the nucleic
 CC acid encoding the polypeptide is useful for preventing, treating, or
 CC ameliorating a medical condition, which involves administering the
 CC polypeptide or the nucleic acid to a mammalian subject. The nucleic acid
 CC is useful for diagnosing a pathological condition or a susceptibility to
 CC a pathological condition in a subject, which involves determining the
 CC presence or absence of a mutation in the nucleic acid, and diagnosing a
 CC pathological condition or susceptibility to a pathological condition
 CC based on the presence or absence of the mutation. The polypeptide, the
 CC nucleic acid and an antibody to the polypeptide are useful for treating
 CC autoimmune disease, Parkinson's disease, silicosis, gastrointestinal
 CC disease, atherosclerosis, haemophilia, thrombocytopenia. The polypeptide,
 CC the nucleic acid and the antibody are useful as immunosuppressive agents,
 CC as adjuvants to enhance immune responses, and as agents to induce higher
 CC affinity antibodies and increase serum immunoglobulin concentrations. The
 CC present sequence represents cDNA encoding a novel human protein. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format direct from USPTO at
 CC seqdata.uspto.gov/sequence.html?docid=20020168711.
 CC
 XX
 SQ Sequence 1966 BP; 391 A; 726 C; 608 G; 241 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 6,91e-43 Length: 1966
 Score: 831.50 Matches: 218
 Percent Similarity: 51.74% Conservative: 49
 Best Local Similarity: 42.25% Mismatches: 142
 Query Match: 33.27% Indels: 107
 DB: 10 Gaps: 19
 US-09-993-966-7 (1-470) x ADB93295 (1-1966)
 QY 1 MetGlyLysLeuHisSerIysProAlaIaValCys---LysArgArgGlnSerProGlu 19
 Db 252 ATGGGAAACTGACGTGAGCAGCGCGCCGCCGCCGCAAGGAGAGAGGCGGAA 311
 QY 20 -----GlyAspSerPheAlaValSerAla---Ala 28

Db 312 GGTTCGCCGCGCTCCGCCGCCCGGACCGGAGGAGAGCTTCGTGGCGTCCGCTACGCG 371
 QY 29 TrpAlaArgLysGlyIleGluGluTrpIleGlyArg-----GlnArgCysPro 44
Db 372 AGCGCCGCAAGAGCGCGGAGAGCGGAGCGGCGCGCGGAGCAAGCAGAGCTGCC 431			
QY 45 GlyIly-----ValSerGlyProArgGlnLeuArgLysIleGlyArgSer 62			
Db 432 AATCGGAGACCCCAAGAGAGGCGCTTCCGG----- 461			
QY 63 ThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGluAsp 82			
Db 462 -----GAGACGAG 470			
QY 83 PheArgLeuGluValAlaLeuProProGluValThrAspGly----- 96			
Db 471 TGTCCCTACAGAGGAGGACTCCCGCTGAAAGAGCTGAGGAGCGCGAGCA 530			
QY 97 ---LeuGlySerGlyAsp---GluLysLysMetGluArgValSerGluProCysProGly 114			
Db 531 CTCCTACGCGCAGATGACGAGAGAGGAGGAGCAAAACCGGAGGCGCGGAGACCGGCG 590			
QY 115 SerLysLysGluLeuLysPheGluGluLeuGlnCysAspValSerMetGluGluAspSer 134			
Db 591 GGCACAGCGC---CTCAACATTGACGACCTCCAGTGGCATCTCTCGGTGAGGAGACGAC 647			
QY 135 ArgGluGluTrpThrPheThrLeuValAspPheAspAsnAsnGlyValValThrArgGlu 154			
Db 648 CGCCAGAGAGGAGAGCTTCACGCTCTGACTTGACACTGCGGAGAGGATCCAGGAG 707			
QY 155 AspIleThrSerLeuLeuHisThrIleTyrGluValValAspSerValAsnHisSer 174			
Db 708 GACATGCCAGCTCATGACACACATCATGAGGTGTGATGCTCGTCAACCATCC 767			
QY 175 ProThrSerSerLysMetLeuArgValLysLeuThrValAlaProAspGlySerGlnSer 194			
Db 768 TCGGCGCAGCAGAAAGCCCTCGGTGAACTCAACGCTCAGCCCTGAGCCCTCAGCAGAG 827			
QY 195 LysArgSerValLeuValAsnGlnAlaAspLeuGlnSerValAspProArgAlaGluThr 214			
Db 828 AGAAGAGAGGAGTCTCTGCTGCGCAGAGCGGAGCGCACCCGTCAGAGATGAGAGGT 887			
QY 215 LysProThrGluAspLeuAspLysSerTrpGluLysGlnArgAlaProLeuArgPheGln 234			
Db 888 GAACGCGCAGAGAGGACCAAGGAGGCGTCAAGAGAGGTGTCTCACACGTCAGAGAGGCC 947			
QY 235 GlyAspSerArgLeuGluGlnSerGlyCysTyrThrHisCysValAspGluAsnIleGlu 254			
Db 948 AGTACTGACCCCGACGCTGCTGAGCGGAGCGGCGCTCACTGCGTGAACGAGAACGAG 1007			
QY 255 ArgArgAsnHisIleTyrLeuAspLeuAlaGlyIleGluAsnTyrThrSerGlnPheGlyPro 274			
Db 1008 CGGAGAAACCACTACCTGACCTGCGCGGAGATTGAGACTACAGTCCCAATTCGAGCCCT 1067			
QY 275 GlySerProSerValAlaGlnLysSerGluLeuProProArgThrSerAsn---ProThr 293			
Db 1068 GGGTCCCTCTCTGTGCAAGCAAGAGAGAGCCCGGAGGAGGAGGCTGCACTCCAGGCC 1127			
QY 294 ArgSerArgSerHisGluProGluAlaIleHisIleProHisArgGlyAspProGln----- 311			
Db 1128 CGGTCCCGCTCCCGAGGAGGAGATACACATGCGCTACACACCGAGTCAAGGTGCTG 1187			
QY 312 -----GlyValAspProAlaSerPheHisPheLeuAspThrProIleAlaLysValSer 329			
Db 1188 GTGGACACATCGTGGTCAAGCTCG-----GAGCTGCTGCTCGGAGCCCTG 1232			
QY 330 GluLeuGlnLysArgLeuArgGlyLysGlnAspGlySerLysHisPheValArgSerPro 349			
Db 1233 GACACGAGCGCCCGCGGAGGAGGCGGAG-----AAGCATTCCTCAAGTCCGCC 1283			
QY 350 LysAlaGlnLysLysSerValGlyValGlyHisValAlaArgGlyAlaArgAsnLysPro 369			
 Db 1284 AAGGCGTCCGGAAGCGC----- 1301

ADQ22400
 ID ADQ22400 standard; DNA; 1959 BP.
 XX
 AC ADQ22400;
 XX
 DT 26-AUG-2004 (first entry)
 XX
 DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 5220.
 XX
 KM soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
 KW ds.
 XX
 OS Homo sapiens.
 XX
 PN WO2004048938-A2.
 XX
 PD 10-JUN-2004.
 XX
 PF 26-NOV-2003; 2003WO-US038193.
 XX
 PR 26-NOV-2002; 2002US-0429739P.
 XX
 PA (PROT-) PROTEIN DESIGN LABS INC.
 XX
 PI Aziz N, Ginsburg WM, Zlotnick A;
 XX
 DR WPI, 2004-441208/41.
 XX
 PT Early detection of soft tissue sarcoma comprises determining expression
 PT of a gene in a first soft tissue sample and a normal soft tissue sample
 PT and comparing the gene expression, also useful in treating soft tissue
 PT sarcoma.
 XX
 PS Example 2; SEQ ID NO 5220; 210bp; English.
 XX
 CC The invention relates to a novel method for detecting soft tissue sarcoma
 CC which comprises obtaining a first soft tissue sample from an individual
 CC and a normal soft tissue sample from the same or different individual,
 CC determining the expression of a gene in both samples and comparing the
 CC expression of the gene in both soft tissue samples, where a higher level
 CC of protein expression in the first soft tissue sample indicates the
 CC presence of soft tissue sarcoma. The method of the invention has
 CC cytostatic applications and may be useful for detecting soft tissue
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC DNA of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.
 XX
 SQ Sequence 1959 BP; 408 A; 677 C; 612 G; 244 T; 0 U; 18 Other;
 Alignment Scores:
 Pred. No.: 4.84e-40 Length: 1959
 Score: 786.00 Matches: 215
 Percent Similarity: 47.24% Conservative: 50
 Best Local Similarity: 38.32% Mismatches: 152
 Query Match: 31.45% Indels: 145
 DB: 12 Gaps: 20
 US-09-993-966-7 (1-470) x ADQ22400 (1-1959)
 QY 1 MetGlyLysLeuHisSerLysProAlaAlaValCys--LYSARGARGLUSerProGlu 19
 DB 98 ATGGGAAATGCACTCGAACACAGCCGCCGCCGCCGCAAGCGAAGAGAGCCCGAA 157
 QY 20 GlyAspSerPheAlaValSerAla---AlaTrpAlaArgLysGlyIleGluGluTrpIle 38
 DB 158 GGGGACACCTTCGTGGGCTCCGCGTACCGAGCGCGCCGCAAGCGCGGAGAGCGGAG 217
 QY 39 GlyArg-----GlnArgCysProGlyIly-----ValSerGlyProArgGln 52
 DB 218 CGGCGCGCGCGGAGCAAGCAGAGAGCTGCCAATGGGAGACCCCAAGAGAGGCGCTTTCCGG 277

QY 53 LeuArgLeuAlaGlyThrIleGlyArgSerThrArgGluLeuValGlyAspValLeuArg 72
 DB 277 ----- 277
 QY 73 AspThrLeuSerGluGluGluAspPheArgLeuGluValAlaLeuProProGlu 92
 DB 278 -----GAGGACCACTGTCCCTACAGGTGGCTCATCCCGCTGAG 316
 QY 93 LysThrAspGly-----LeuGlySerGlyAsp---GluYbLys 104
 DB 317 AAAGCTGAGGGCCCGAGACCCCGGAGCAACTCTTACAGCCCAATATGAGAGAGGCA 376
 QY 105 MetGluArgValSerGluProCysProGlySerLysLeuGluLeuLysPheGluLeu 124
 DB 377 GCAACCGGAGGGCCCGGAGAGACCGGGGCGGACAGCC---CTCAACATTAGCAGCACTC 433
 QY 125 GlnCysAspValSerMetGluGluAspSerArgGlnGluTrpThrPheTrpLeuTyAsp 144
 DB 434 CAGTGCATGTCTCGGTGGAGGAGAGAGACCCGAGAGAGTGAAGTTCACAGCTCTATGAC 493
 QY 145 PheAspAsnAsnGlyLysValThrArgGluAspIleThrSerLeuLeuHisThrIleTy 164
 DB 494 TTGACAACTGCGGAAAGGTACAGAGAGACATGTCCAGCTCATGACACATCTAT 553
 QY 165 GluValValAspSerSerValAsnHisSerProThrSerSerLysMetLeuArgValLys 184
 DB 554 GAGGTGCGTATGCTCTCGTCAACCACTCTCGGCGAGCAGCAAGACCTCCGTGTGAAG 613
 QY 185 LeuThrValAlaProAspGlySerGlnSerLysArgSerValLeuValAsnGlnAlaAsp 204
 DB 614 CTAAACGTCAGCCCTGAGCCCTCAGACAGAGAGAGAGGCTCTCTGCTGCGCAGGAC 673
 QY 205 LeuGlnSerLysLysProAlaGluThrLysProThrGluAspLeuArgSerTrpGlu 224
 DB 674 CGGAGCCCAACCCGTGTGACAGAGTGAAGGCTGAACCTGGCAGAGACCAAGCGGTGAC 733
 QY 225 LysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGluGlnSerGlyCys 244
 DB 734 AGGAGTGTCTGTGACACGTCAAGAGGCCAGATACGACCCCAAGCCCTCTCGAGCGG 793
 QY 245 TyrHisIleCysValAspGluAsnIleGluArgArgAsnHisTyrLeuAspLeuAlaGly 264
 DB 794 GGCCCTCACTGCTGTGACGAGAACACGAGACCGAGAAACCATCACTGCGACCTCGCGGG 853
 QY 265 IleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValAlaGlnLysSerGlu 284
 DB 854 ATTGAGAACTACACGTTCAGATTGCGCTGGGTCTCAGCTGTGCGAAGAGAGAGCTCC 913
 QY 285 LeuPro----- 286
 DB 914 GCTCCAGACACACAGTGGGAGCAAGGCTAAGAGAGTGGCCTTTGACGAGAGCTGTGG 973
 QY 286 ----- 286
 DB 974 AGCCAGGACAGGTACACCAAGTGGCCAGGCCCTTCCCTTCAAGGGCTGTGGCCGTGA 1033
 QY 286 ----- 286
 DB 1034 CTGACAGCTTGCTTAACAGACTGGCTCAGGGTCCCTCTGTGCAAGCAAGAGGAGC 1093
 QY 287 -----ProArgThrSerAsnProThrArgSerArgSerHisGluProGluAla 302
 DB 1094 CCAAGGACAGGCTCTGCACTTCCAGGCCCG---TCCGCTCTCCAGAGCCAGATACA 1149
 QY 303 IleHisIleProHisArgLysProGln-----GlyValAspProAlaSerPhe 318
 DB 1150 CATGCCGTACACACCGCAGGTCAAGGTCTGTGTGAACACATCGGCGAGGCTCG--- 1206
 QY 319 HisPheLeuAspThrProIleAlaLysValSerGluLeuGlnGlnArgLeuArgGlyThr 338
 DB 1207 -----GAGCTGTGCGCGGCGCTGGACCGAGCCCGCGGCGGCGGCGG 1254
 QY 339 GlnAspGlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyVal 358

```
Db      1255 GAG-----AAGCAGTTCCTCAAGTCCCAAGGGCTCCGGGAAAGCCGCTGGGNNN 1305
QY      359 GlyHisValAlaArgGlyAlaArgAsnLysProLeuGlyProAlaIleProAlaVal 378
Db      1306 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1356
QY      379 SerPro-----SerAlaHisLeuAlaAlaSerProAlaLeuLeuPro 392
Db      1357 CTGCGCGCCCAAGGCCCCCTCAGAGCGGCCACCACTCCCGCAGCCCCCACCAG----- 1407
QY      393 SerLeuAlaProLeuGlyHisLysLysHisLysAlaLysGlyLysGlyLysGlyLysGly 412
Db      1408 -----CCACCTTACGGGCGCACAGCGGTACCGCCAAAGGGCAGGAG-----GGC 1452
QY      413 CysArgGlyLeuGlnAlaProLeuAlaSerGlyLysProValLeuGlyArgGlyHis--- 431
Db      1453 CACTCGCCTCAAGGCCCCCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 1503
QY      432 ---LeuArgGlyLeuProAlaLeuValValTyrGlySerGlnAlaGlyGlnProValGln 450
Db      1504 GTGGTGGGAGCTGCGCCGACCGCCAGCAGAGAGAGGAGTACGCGGTGCCAGTATCCAG 1563
QY      451 ArgHisGlyHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 467
Db      1564 CGGCACAGACACACACACACACACACACACACACACACACACACACACACACACACACCTTC 1623
QY      468 Tyr 468
Db      1624 CAC 1626
```

Search completed: December 30, 2004, 04:02:31
Job time : 874 secs

THIS PAGE IS BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 30, 2004, 00:53:19 ; Search time 8160 Seconds
(without alignments)
2723.794 Million cell updates/sec

Title: US-09-993-966-7
Perfect score: 2499
Sequence: 1 MGKLSKPAACVCKRRSPG.....RHHHHHHHHHHHHFYOT 470

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues.

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xlp
-Q/cgn2_1/USPTO.spool.p/US09993966/runat_29122004_130758_8996/app.query.fasta_1.647
-DB=GenEmbl -OFMT=fasta -SUFFIX=rge -MINMATCH=0.1 -LOOPT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09993966 @CEN_1_1_7406 @runat_29122004_130758_8996 -NCPU=6 -ICPU=3
-NO NMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl: *
1: gb_ba: *
2: gb_hcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_scs: *
12: gb_ey: *
13: gb_un: *
14: gb_vi: *

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	2499	100.0	1438	9 AF358135 Homo sapi
2	2499	100.0	1786	9 AY061883 Homo sapi
3	2499	100.0	1788	9 AB062886 Homo sapi
4	2499	100.0	2604	9 BC051288 Homo sapi

Result 1	AF358135	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	TITLE	JOURNAL	FEATURES	source
5	2493	99.8	1438	6	AR405917														
6	2192.5	87.7	1416	10	AF343352														
7	2191.5	87.7	1765	10	BC034838														
8	2189.5	87.6	1671	10	AF358134														
9	2174.5	87.0	1401	6	AR405916														
10	2167.5	86.7	1431	6	AX172351														
11	1677	67.1	1053	6	C0721681														
12	1210.5	48.4	207486	2	AC145175														
13	1191.5	47.7	163319	9	HS403140														
14	1191.5	47.7	169773	9	AC007608														
15	1191.5	47.7	170791	9	AC007334														
16	1191.5	47.7	178790	2	AC145238														
17	1135	45.4	194869	2	AC145003														
18	1135	45.4	198189	2	AC145043														
19	1112	44.5	232509	2	AC134093														
20	1076.5	43.1	208061	2	AC145254														
21	1075	43.0	240965	2	AC126054														
22	1055.5	42.2	199277	2	AC145445														
23	1029	41.2	139145	2	AC146326														
24	939	37.6	181401	2	AC146663														
25	872.5	34.9	1967	10	BC019952														
26	864.5	34.6	1947	10	AF358136														
27	846.5	33.9	1682	9	AF358137														
28	846.5	33.9	1798	9	AB062887														
29	793	31.7	1285	6	BC012176														
30	780.5	31.2	1885	6	AX780854														
31	641.5	25.7	1309	6	BC004940														
32	637.5	25.5	212531	2	AC145178														
33	628	25.1	1191	6	C0715890														
34	583.5	23.3	1307	6	AR405919														
35	314.5	12.6	181105	5	BX664719														
36	285.5	11.4	99395	2	AC010446														
37	272	10.9	162641	2	AC141571														
38	272	10.9	208878	2	AC141571														
39	272	10.9	162641	2	AC141571														
40	272	10.9	246241	2	AC135365														
41	246	9.8	1950	6	AX557538														
42	240.5	9.6	2340	6	C0741468														
43	224	9.0	237332	2	AC131846														
44	215.5	8.6	2691	6	C0605931														

ALIGNMENTS

RESULT 1
AF358135
LOCUS Homo sapiens naked cuticle-1 (NCKD) mRNA, complete cds.
DEFINITION
ACCESSION AF358135
VERSION AF358135.1 GI:14211713
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1438)
Wharton,K.A., Jr., Zimmermann,G., Rouselet,R. and Scott,M.P.
Vertebrate proteins related to Drosophila Naked Cuticle bind
dishevelled and antagonize Wnt signaling
Dev. Biol. 234 (1), 93-106 (2001)
JOURNAL
MEDLINE
2125530
PUBMED
11356022
REFERENCE
2 (bases 1 to 1438)
Wharton,K.A., Jr., Zimmermann,G. and Scott,M.P.
Direct Submision
Submitted (08-MAR-2001) Pathology and Molecular Biology, University
of Texas Southwestern Medical School, 5323 Harry Hines Blvd.,
Dallas, TX 75390-9072, USA
Location/Qualifiers
1..1438
/organism="Homo sapiens"
/mol_type="mRNA"

```

gene
/ db_xref="taxon:9606"
/ chromosome="16"
/ map="16q12"
1..1438
/ gene="NKD1"
8..1420
/ gene="NKD1"
/ note="contains EFX domain, similar to Drosophila melanogaster naked cuticle (nkd) and to Mus musculus Nkd1"
/ codon_start=1
/ product="naked cuticle-1"
/ protein_id="AAK57484.1"
/ db_xref="GI:14211714"
/ translation="MGKLSKPAVACRRSPGSDSPASAAARKIEEWIGRORCP
GYSGRPLAIGTGRSTRELVDVLRDPLSRERDPRFLVALPEKDTGSGDPE
KKRRVSEPCRSKKQKLFEBLQCDVSMEDSROBEFTLYTDPDNNEKTRBETISLL
HTLYEVDSSVNSPTSSKMLRYKLIVAPDSQSKSVLVNQADLQASAPAEKPEPE
DLSEWKKORAPLRFGDSRLBQSCYHNCVDENIERNNHYLDLAGIENTYSGPDS
PSVAQSESLPRTSNPTSRSHSEPAIHIHRRKQGVDPASPFHLDTPAKVELOQR
LRGTODGSKHFRVSPKAGSKSVGVHARARANKPPLGPAPVPSFAHLAASPLLP
SLAPLGKHKHRAKESQOGCRGLQAPLASGPVLGRHLRELPLVTVYESQAGPVG
RHHHHHHHHHHHHHPTQT"

ORIGIN

Alignment Scores:
Pred. No.:      3,17e-112      Length:      1438
Score:          2499.00      Matches:      470
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
DB:              9          Gaps:      0

US-09-993-966-7 (1-470) x AF358135 (1-1438)

QY      1 MetGlyLeuHisSerIysProAlaAlaValCysAlaArgArgIuseProGluGly 20
Db      8 ATGGGGAACCTTCACTCCAAAGCCGGCCGCTGTGCAAGCGCAGAGAGAGCCGGAAGGT 67
QY      21 ArgSerPheAlaValSerAlaAlaTTPAlaArgIysGlyIleGluGluTTPIlleGlyArg 40
Db      68 GAAGAGCTTCGCCGTGAGCGCTGCTGGGCTCGAAGAGGCATCGAGAGTGGATGGAGAGA 127
QY      41 GluArgCysProGlyGlyValSerGlyProArgGluLeuArgLeuAlaGlyThrIleGly 60
Db      128 CAGCGCGCCCGGGCGGTGTCTCGGGACCCCGACAGCTCGCGTTGGCGGGACCATAGGC 187
QY      61 ArgSerThrArgGluLeuValGlyAspValLeuArgPheThrLeuSerGluGluGlu 80
Db      188 CGAAGCACCCGGGAGCTCGTGGGCGACGTGTTGAGAGACAGCGCTCAGCGAGGAAGAGAG 247
QY      81 AspAspPheArgLeuGluValAlaLeuProProGluLysThrArgGlyLeuGlySerGly 100
Db      248 GAGGACTTTGGCTGGAGAGTGGCCCTGCTCTGAGAGAGACTGACGGGCTGGGCGAGCGGA 307
QY      101 AspGluLysIysMetGluArgValSerGluProCysProGlySerIysGluLeuLys 120
Db      308 GATGAGAGAAGATGAGAGAGTGAAGCAACCTTGCCAGGCTCAAGAAAGCACTGAAG 367
QY      121 PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGlnGluTTPThrPhe 140
Db      368 TTGTAAGAGCTCCGAGTGCAGTGTTCATGGAGAGAGACAGCCGGCAGAGTGAACCTTC 427
QY      141 ThrLeuTyrAspPheAspAsnArgGlyLysValThrArgGluAspIleThrSerLeuLeu 160
Db      428 ACCCTGATATACITTTGCAACACGCAAGGTCAACCGAGAGGACATCACCAAGCTTCGTG 487
QY      161 HisThrIleTyrGluValValaAspSerSerValAsnHisSerProThrSerSerIysMet 180
Db      488 CACACCACTATAGGGGTGAGATCTCTGTCAACACTCCCAACATCCAGCAAGATG 547
QY      181 LeuArgValIysLeuThrValAlaProAspGlySerGlnSerIysATGSerValLeuVal 200
Db      548 CTGCGGGTAAAGCTCACCGTGGCCCGCATGTGGACGACAGCAAGAGAGAGTGTCTTGTCTC 607

```

```

QY      201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIysProThrGluAspLeu 220
Db      608 AATCAGGCTCACTCTGCAGAGCGCAAGGCCCCCGAGCAGAGACCAAGCCACTGAGGACTTG 667
QY      221 ArgSerTTPGluLysIysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
Db      668 CGGAGCTGGAGAGAAGAGCAGCGAGCCCGCTCAGGTTCACAGGATGACACCCGCTGGAG 727
QY      241 GluSerGlyCysTyrHisHisCysValAspGluAsnIleGluLysArgAsnHisTyrLeu 260
Db      728 CAGTGTGGCTGTCAACCAATGCGTAGATGAAGAACATCAAGAGAGAAACCACTACTTA 787
QY      261 AspLeuAlaGlyIleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValAla 280
Db      788 GATCTCCCGGGATAGAAACTACACTCCCAATTTGGGCTGGCTCCGTTCCGTGGCC 847
QY      281 GluLysSerGluLeuProProArgTyrSerAsnProThrArgSerArgSerHisGluPro 300
Db      848 CAGAGTCAGAACTGCCCCCGCACCTCCATCCCACTCGATCTCGCTCCCATGAGACCG 907
QY      301 GluAlaIleHisIleProHisArgIysProGlnGlyValAspProAlaSerPheHisPhe 320
Db      908 GAAGCCATCCACATCCACACGAAAGCCCCAAGCGTGAACCGGCTCTTCCACTTC 967
QY      321 LeuAspThrProIleAlaLysValSerGluLeuGlnGluArgLeuArgGlyThrGluAsp 340
Db      968 CTTGACACCCCAATCGCCAGAGTCTCAGAGCTCAGAGACGGCTCGGGGCAACCCAGGAC 1027
QY      341 GlySerLysHisPheValaArgSerProLysAlaGlnGlyLysSerValGlyValHis 360
Db      1028 GGGAGCAAGCACTTTGTGAAGTGTCCCCCAAGGCCAGGCAAGATGTGGTGTGGGCAC 1087
QY      361 ValAlaArgGlyValaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
Db      1088 GTGGCCAGAGGGCAGAAACCAAGCCCCCTGTGGACCCGCATCTCGGTGTCTCCC 1147
QY      381 SerAlaHisLeuAlaHisSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
Db      1148 TCGGCCACTGTGGTGCAGCCCGGCTCTCTCCCTCTTACGCCCTCGGGCACAG 1207
QY      401 LysHisLysHisArgAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
Db      1208 AAGGACAAAGCACGAGGCAAGAGCAGACAGGAGTGGCGGGCTCGAGGCAACCACTG 1267
QY      421 AlaSerGlyGlyProValLeuGlyArgGluHisLeuArgGluLeuProAlaLeuVal 440
Db      1268 GCCTCAAGTGGCCCTGTCTGGGGCGGGAGCACCTGCGGAGCTGCCCCGCTTGTGTG 1327
QY      441 TyrGluSerGlnAlaGlyGlnProValGlnArgHisGluHisHisHisHisHisGluHis 460
Db      1328 TATGAGAGCCAGGCGGGGCGAGCGGTCCAGAGACATAGCACCAACCACTGAACAT 1387
QY      461 HisHisHisTyrHisHisPheTyrGlnThr 470
Db      1388 CACCACCATTAACACCACTTCTACAGACA 1417

RESULT 2
AY061883      1786 bp      mRNA      linear      PRI 26-DEC-2001
LOCUS      Homo sapiens naked protein (NKD) mRNA, complete cds.
DEFINITION      AY061883
ACCESSION      AY061883.1 GI:17978536
VERSION      AY061883.1
KEYWORDS
SOURCE
ORGANISM      Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1786)
Yan,D., Wisemann,M., Rohan,M., Chan,V., Jefferson,A.B., Guo,L.,
Sakamoto,D., Coothern,R.H., Fuller,J.H., Reinhard,C., Garcia,P.D.,
Randazzo,F.M., Escobedo,J., Fanl,W.J. and Williams,L.T.
Elevated expression of axin2 and nkhd mRNA provides evidence that

```


Wnt/beta -catenin signaling is activated in human colon tumors
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 14973-14978 (2001)

JOURNAL
PUBMED
11752446
2 (bases 1 to 1786)
REFERENCE
Rohan, M., Chan, V., Yan, D., Fan, L., W. J. and Williams, L. T.
AUTHORS
TITLE
Direct Submission
JOURNAL
Submitted (05-NOV-2001) Cancer Biology, Chiron Corporation, 4560
Horton Street, Emeryville, CA 94608, USA

FEATURES
Location/Qualifiers

source

1..1786

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

1..1786

/gene="NKD"

159..1571

/gene="NKD"

/note="HNKD"

/product="naked protein"

/protein_id="AAJ32374.1"

/db_xref="GI:17978537"

/translation="MGKLHSPKPAVCSKRESPEGSFVSAWARKGIEEMIGRCP
GVSGRPLAGTIGSTRELVDYRLDRLSEEDPRLVALPEKTDGSGDE
KMERVSEPCRSKKOLKEBELCCDVSMEEDSDEMTFTYDIDNNGKTRDITSL
HTIYEVDSSVNHSPSSKMLRYALYAPDSQSKSVLVNADLDLQARPRATKPE
DLKSWKQDAPLRFGDSRLSQSCYNHCVDENIERNNYLDLAGINTSGFQPS
PSVAKSELRPTSPNPTSRSHPEALIHIRKPOGVDPHFLLDPIKVELQOR
LRGTODGSKHFRVS PRAQGSKGVGVNARANKPFGPAI PAVSPAHLAAPLPL
SLAPLGHKHKHAKESQSGCRGLQAPLAGSVPVLRHRELPAVVYESQAGPQV
RHHNNHHNNHHNNHNYTC"

ORIGIN

Alignment Scores:

Pred. No.: 4,08e-112 Length: 1786
Score: 2499.00 Matches: 470
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-993-966-7 (1-470) x AY061883 (1-1786)

QY 1 MetGlyLeuLeuHisSerLysProAlaAlaValCysLysArgArgGlySerProGlyGly 20
DB 159 ATGGGAACTTCACTCAAGCCGCGCGCTGTGCAAGCGAGGAGAGCCCGAAGGT 218
QY 21 AspSerPheAlaValSerAlaAlaTTPAlaArgLysGlyLeuGlyGlyGlyArg 40
DB 219 GACAGCTTGCCCTGAGCGCTGCGCTGCGAGGAGGAGTCCAGAGAGTGGATCGAGGA 278
QY 41 GlnArgCysProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGly 60
DB 279 CAGCGCTGCGCGCGCGGTGTCTCGGAGCCCGACAGCTGTGGGTGGGACCATAGGC 338
QY 61 ArgSerThrArgGlyLeuValGlyAspValLeuArgAspThrLeuSerGlyGlyGlyGly 80
DB 339 CGAAGACCCGGAGCTCGTGGGCGAGCTGTGAGAGACAGCTCAGCGAGGAGAGAG 398
QY 81 AspAspPheArgLeuGlyValAlaLeuProProGlyLeuThrArgGlyLeuGlySerGly 100
DB 399 GACGACCTTTCGCTGGAAGTGGCTCGCTCTGAGAGAGCTAGCGGCTGGCGAGCGGA 458
QY 101 AspGlyLeuLysMetGlyArgValSerGlyProCysArgProGlySerLysLysGlnLeuLys 120
DB 459 GATGAGAGAGAGTGAAGAGAGTGAAGAGACCTGCGCAGAGCTCCAGAGAGAGCTGAG 518
QY 121 PheGlyGlyLeuGlnCysAspValSerMetGlyGlyLeuAspSerArgGlnGlyThrPhe 140
DB 519 TTGGAAGAGCTCAGAGCGAGCTGTCTCATGAGAGAGAGAGCCGCGAGAGTGAAGCTTC 578
QY 141 ThrLeuArgAspPheAspAsnArgLysValThrArgGlyLeuArgLeuThrSerLeuLeu 160

DB 579 ACCCTGATGACTTTGACAAACGAGAGTCAACCCGAGAGAGCATCAGCTTCTG 638
QY 161 HisThrIleArgGlyValValAspSerSerValAsnHisSerProThrSerSerLysMet 180
DB 639 CACACCATCATGAGGTGGTGAAGCTCTGTCTCAACCATCCCAATCATCAGAGATG 698
QY 181 LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysArgSerValLeuVal 200
DB 699 CTGCGGGTAAAGCTCACCTGTGGCCCGCATGGAGCGCAGAGAGAGAGAGAGCTTGTTC 758
QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGlyThrLysProThrGlyAspLeu 220
DB 759 AATCAGCTGACCTGCGAGAGCGAGAGCCCGGAGAGAGAGAGAGAGAGAGAGAGAG 818
QY 221 ArgSerTrpGlyLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGly 240
DB 819 CGAGAGTGGAG 878
QY 241 GlnSerGlyCysArgLysHisCysValAspGlyLeuGlnGlyArgGlnArgLysLeu 260
DB 879 CAGCTGCTGCTGCTACCACTTCGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 938
QY 261 AspLeuAlaGlyLysLeuArgLysThrSerGlnPheGlyProGlySerProSerValAla 280
DB 939 GATCTGCGCGGATGAG 998
QY 281 GlnLysSerGlyLeuProProArgLysSerAsnProThrArgSerArgSerHisGlyPro 300
DB 999 CAGAGTCAAGATGAG 1058
QY 301 GlnAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPhe 320
DB 1059 GAGAGCATTCACATCCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1118
QY 321 LeuAspThrProIleAlaLysValSerGlyLeuGlnGlnArgLeuArgGlyLysThrGlnAsp 340
DB 1119 CTGACACCCCAATCCGCAAGTCTCAGAGCTCAGAGAGAGAGAGAGAGAGAGAGAGAG 1178
QY 341 GlySerLysHisAspValArgSerProLysAlaGlnGlyLysSerValGlyValHis 360
DB 1179 GAGAGCAAGCATTTGTGAGGTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1238
QY 361 ValAlaArgGlyValArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
DB 1239 GTGGCCAG 1298
QY 381 SerAlaHisLeuAlaLysProAlaLeuProSerLeuAlaProLeuGlyHisLys 400
DB 1299 TCCGCCCACTGGCTGCGAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1358
QY 401 LysHisLysHisArgAlaLysGlySerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
DB 1359 AAGCACAAG 1418
QY 421 AlaSerGlyGlyProValLeuGlyArgGlyHisLeuArgGlyLeuProAlaLeuVal 440
DB 1419 GCCTCAGTGGCCCTGTCTGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1478
QY 441 TyrGlnSerGlnAlaGlyGlnProValGlnArgHisGlnHisHisHisHisHisHis 460
DB 1479 TATGAG 1538
QY 461 HisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 470
DB 1539 CACCAACATTACCAACATTCTACAGAGAG 1568

RESULT 3
AB062886 1788 bp mRNA linear PRI 20-OCT-2001
LOCUS Homo sapiens mRNA for Dvl-binding protein NKD1, complete cds.
DEFINITION
AB062886
ACCESSION
AB062886.1 GI:16303259
KEYWORDS

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1
TITLE Molecular cloning, gene structure, and expression analyses of NKD1 and NKD2
JOURNAL Int. J. Oncol. 19 (5), 963-969 (2001)
MEDLINE 21490203
PUBMED 11604995
REFERENCE 2 (bases 1 to 1788)
AUTHORS Katoh, M.
TITLE Direct Submission
JOURNAL Submitted (08-JUN-2001) Masaru Katoh, National Cancer Center Research Institute, Genetics and Cell Biology Section, Tsukiji 5-chome, Chuo-Ku, Tokyo 104-0045, Japan (E-mail: mkatoh@ncc.go.jp, Tel: 81-3-3542-2511)
FEATURES
source location/Qualifiers
1..1788
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="human fetal tissues"
1..1788
/gene="NKD1"
6..1418
/gene="NKD1"
/codon_start=1
/product="Dvl-binding protein NKD1"
/protein_id="BAB70500.1"
/db_xref="GI:16103260"
/translation="MGKLSKPAVCKRSPGDSFVSAAMARKIEEMIGRQRP
GVSQPRQLAGTIGKSTRELVGDVLRDTLSEBEDPFLVLPPEKTDGLSGSD
KKNRVEPCPGSKQLEKFELOCDVMSBDSRQEWFTLYDPMNKVTRIDITSL
HTIYEVDSSVNHSPISKMLRYLTVAPGSGSKSVLYNOADLSARPAETKPE
DLRWSEKKORAPLRPGDSRLBOSGCHNCVDENIERNHYLDLAGIENTSGRPS
PSYAKSELAPRTSNPRSRSHPEAITHIRKRGQVDPASPHLDTPIKVSRLQOR
LRGTODGSKRFVSPKQKSGVGVGVARAKARKNRPRIAPVSPSNHLAASALLP
SLAPLGHKKHAKESQCCRGDLAPLAGGVPVLRGKHLRELVALVVSQKQPVQ
RHHNHNNHHNHNHYNYQ"

ORIGIN

Alignment Scores:

Pred. No.: 4,09e-112 Length: 1788
Score: 2499.00 Matches: 470
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-993-966-7 (1-470) x AB062886 (1-1788)

QY 1 MetGlyLeuSerLeuHisSerLeuProAlaAlaValCysLeuSarGArgGluSerProGluGly 20
DB 6 ATGGGGAAGAACTTCACTCCAGCCGCGCGCGCTGTGCAAGCCGCAAGGAGCGCGGAAGGT 65
QY 21 AspSerPheAlaValSerAlaAlaTTPAlaArgLysGlyTllegluGluTTPAlaGlyArg 40
DB 66 GACAGCTTCGCGCGGACGCGCTGCGGAGCTCGGAAGGCGATCGAGAGTGAGTGGGAGA 125
QY 41 GluArgCysProGluGlyValSerGlyProArgGluLeuArgLeuAlaGlyThrTllegly 60
DB 126 CAGCGCTGCGCGGCGGAGTGTCTCGGAGCCCGGACAGCTGCGGTTGGCGGCGACCATAGGC 185
QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGluGlu 80
DB 186 CGAAGACACCGGAGGAGCTGTGGGCGACGCTGTGAGAGACAGCTCAAGCGAAGAGAGAG 245
QY 81 AspAspPheArgLeuGluValAlaLeuProGluLysThrArgGlyLeuGlySerGly 100
DB 246 GACAGCTTCGCGCTGGAAGTGGCGCTGCTCTGAGGAAGACTGACGAGCTGGGCGAGCGGA 305

QY 101 AspGluLeuSerMetGluArgValSerGluProCysProGlySerLysGluLeuLys 120
DB 306 GATGAGAGAAAGATTGAGAGAGTGAAGAACCTTCCCAAGCTCCAAAGAGAGCTGAAG 365
QY 121 PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGlnGluTTPThrPhe 140
DB 366 TTGAAGAGCTCCAGTCCGACGCTGTCCATGAGAGAGACAGCCGCGAGGTGAGACTTC 425
QY 141 ThrLeuThrAspPheAspPheAsnGlyLysValThrArgGluAspPheThrSerLeuLeu 160
DB 426 ACCCTGATGACTTTGACCAACACGCGAAGGTACCCGAGAGACATCACAGCTTGCTG 485
QY 161 HisThrTlleglyArgGluValAlaAspSerSerValAsnHisSerProThrSerSerLysMet 180
DB 486 CACACCATCTATAGAGGTGGTGACTCTCTGTCAACACCTCCCAATCATCAGAGATG 545
QY 181 LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysArgSerValLeuVal 200
DB 546 CTCGGGGTAAAGCTCACCCGTGGCCCCGATGGCAGCCAGAGCAAGAGAGAGCTTCCTGTC 605
QY 201 AsnGluAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeu 220
DB 606 AATCAGGCTGACCTGACAGCGGAGGCGCCGAGCAGAGACCAAGCCACTGAGAGACTTG 665
QY 221 ArgSerTPGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
DB 666 CGGAGCTGGGAGAAAGAAAGACGACGAGCCCGCTCAGGTTCCAGGTTGACACCCGCTGAG 725
QY 241 GlnSerGlyCysTyrrHisHisCysValAspGluAsnTllegluArgArgAsnHisTyrrLeu 260
DB 726 CAGTCTGGCTGTACCAACCAATTCGATGAGAAACATCAAGAGAGAAACCACTACTTA 785
QY 261 AspLeuAlaGlyTllegluAsnTyrrThrSerGlnPheGlyProGlySerProSerValAla 280
DB 786 GATCTGCCCGGATAGAAACCTACAGTCCCAATTTGGAGCTTGCTCCCTTCCTGAGCC 845
QY 281 GlnLysSerGluLeuProProArgThrSerAsnProThrArgSerArgSerHisGluPro 300
DB 846 CAGAACTCAGAACTGCCCCCCCCCAGCTCCCAATCCCACTGATCTGCTCCATGAGACCG 905
QY 301 GluAlaTlleglHisTleProHisArgGlySerProGlnGlyValAspProAlaSerPheHisPhe 320
DB 906 GAAGCCATCCATCCACACACGAAAGCCCCAAGGCGCTGAGACCCGCGCTCTCCACTTC 965
QY 321 LeuAspThrProTllegluAlaLysValSerGluLeuGlnGlnArgLeuArgGlyThrGluAsp 340
DB 966 CTTGACACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAACGGCTCCGAGGACCCACAGAC 1025
QY 341 GlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValGlyHis 360
DB 1026 GGAAGACAGCACTTTGTGAGGTCCCCCAAGGCCCAAGGATGTGGGTGGGCAC 1085
QY 361 ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaTllegluAlaValSerPro 380
DB 1086 GTGGCCAGAGGGGCAAGAAACAAAGCCCCCTGTGGACCCCGCAATCCCTGGGTGTCCCC 1145
QY 381 SerAlaHisLeuAlaAspProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
DB 1146 TCGGCCCACTGGGTGCAAGCCCGGCGCTCTCTCCCTTCCAGCCCCCTTCGGGACAGAG 1205
QY 401 LysHisLysHisArgAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
DB 1206 AAGCACAAGACAGGAGCCAGAGAGCAAGAGCAAGGCTGCCGGGCGCTGAGGACCACTG 1265
QY 421 AlaSerGlyGlyProValLeuGlyArgGluHisLeuArgGluLeuProAlaLeuValVal 440
DB 1266 GCTTCAGGTGAGCTGCTGTGGGCGGAGCACTGCGGAGAGCTGCGCGCTTGTGTGTG 1325
QY 441 TyrrGluSerGlnAlaGlyLysProValGlnArgHisGluLysHisHisHisHisHisHis 460
DB 1326 TATGAGAGCCAGGCGGGGAGCCGGTCCAGAGCATATGAGACCAACCAATGAAACAT 1385
QY 461 HisHisHisTyrrHisHisPheTyrrGlnThr 470

Db 1386 CACCACATTACCACTTCTACAGACA 1415

RESULT 4
BC051288

LOCUS
DEFINITION Homo sapiens naked cuticle homolog 1 (Prosochilla), mRNA (cDNA clone MGC:59666 IMAGE:6646089), complete cds.

ACCESSION BC051288

VERSION BC051288.1 GI:30410965

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE
AUTHORS Krausner, R.D., Collins, F.S., Wagner, L.H., Derge, J.G., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhac, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Szelest, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Uedlin, T.B., Toohy, Yuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mulhany, S.J., Boeak, S.A., McEwen, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.R., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Sanchez, J., Helton, E., Kettman, M., Madan, A.C., Shewchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marz, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 2604)

Strausberg, R.

Direct Submission

Submitted (25-APR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc_mgc@hgti.nih.gov

Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakeley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R., Medoro, Q.L., Maiello, C., Maskeri, B., Mastrian, S.D., McCluskey, J.C., McDowell, J., Pearson, R., Stantiridop, S., Thomas, P.J., Touchman, J.W., Young, A., Vogt, J.L., Walker, M.A., Wecherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>

Series: IRAL Plate: 47 Row: 0 Column: 23

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 20127660.

Location/Qualifiers

1..2604

/organism="Homo sapiens"

gene

CDS

ORIGIN

Alignment Scores:

Pred. No.: 6,35e-112 Length: 2604

Score: 2499.00 Matches: 470

Percent Similarity: 100.00% Conservative: 0

Best local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 9 Gaps: 0

US-09-993-966-7 (1-470) x BC051288 (1-2604)

QY 1 MetGlyLeuLeuHisSerLysProAlaAlaValCylValYbArGArgJuserProGluGly 20

DB 225 ATGGGAAATCTTCACTCTCAAGCCGCGCGGTGTGCAAGGCGAGAGAGAGAGCT 284

QY 21 AspSerPheAlaValSerAlaAlaATPAlaArgLysGly11egluGluTTP11eglyArg 40

DB 285 GACAGCTTCGCGCGAGCGCTGCTGCGGAGGAGGATTCAGAGAGTGGATCGGAGGA 344

QY 41 GlnArgCysProGlyGlyValSerGlyProArgGluLeuArgLeuAlaGlyThr11egly 60

DB 345 CAGCGCTGCCCGGCGGCGGTGTCTCGGAGCCCGACAGCTGCGGTGGCGGACCATAGGC 404

QY 61 ArgSerThrArgGluLeuValAlaLeuProProGluTyrThrArgGlyLeuGlySerGly 80

DB 405 GAAAGCACCCGGAGGCTGTGGGAGAGTGTGAGACACGCTCAGCGAGAGAGAGAG 464

QY 81 AspAspPheArgLeuGluValAlaAlaLeuProProGluTyrThrArgGlyLeuGlySerGly 100

DB 465 GAGGACTTTCGGCTGAGAGAGGCGCTGCTCCTCGAAGAGCTAGCGGCTGGGAGAGGGA 524

QY 101 ArgGluLysLysMetGluValSerGluProCysProGlySerLysLysLysLysLys 120

DB 525 GATGAGAAAGAGTGAAG 584

QY 121 PheGluGluLeuGluGlnCysAspValSerMetGluGluAspSerArgGlnGluTTPThrPhe 140

DB 585 TTGGAAGAGCTTCAGTGCAGCGTGTCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAG 644

QY 141 ThrLeuTyrAspPheAspAsnAnglyLysValThrArgGluAspLysLysLysLysLys 160

DB 645 ACCCTGTATGACTTTGACAAACAGGAGAGAGTACCCGAGAGAGAGAGAGAGAGAGAG 704

QY 161 HisThrIleTyrGluValValAspSerSerValAsnHisSerProThrSerSerLysMet 180

Db 705 CACACCATCTATGAGGGTGGAGCTCTCTGTCAACCACTCCCAATCCAGCAATG 764
Qy 181 LeuArgValIysLeuThrValAlaProAspGlySerGlnSerIlybArgSerValLeuVal 200
Db 765 CTGGGGGTAAAGCTCACCGTGGCCCCCGATGGCAGCCAGCAAGAGAGAGGCTCTTGTCTC 824
Qy 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIlybProThrGluAspLeu 220
Db 825 AATCAGGCTACCTGACAGAGCCGAGGAGCCGAGCAGACCAAGCCCACTGAGAGACTTG 884
Qy 221 ArgSerTrpGluIlybIlybGlnArgAlaProLeuArgPheGlnIlybSerArgLeuGlu 240
Db 885 CGGAGCTGGAG 944
Qy 241 GlnSerGlyCysTrpThrIlybIlybValAspGluAsnIlybGluArgAsnIlybTrpLeu 260
Db 945 CAGTCTGCTGTACCACTTGGTATGAGAACATCGAGAGAGAGAGAGAGAGAGAGAGAG 1004
Qy 261 AspLeuAlaGlyIlybGluAsnTrpThrSerGlnPheGlyProGlySerProSerValAla 280
Db 1005 GATCTCGCCGGAG 1064
Qy 281 GlnIlybSerGluLeuProArgTrpSerAsnProThrArgSerArgSerIlybGluPro 300
Db 1065 CAGAGCTCAG 1124
Qy 301 GlnAlaIlybIlybIlybProIlybArgIlybProGlnIlybValAspProAlaSerPheIlyb 320
Db 1125 GAGGCTCAG 1184
Qy 321 LeuAspThrProIlybAlaIlybValSerGlnLeuGlnIlybArgValGlyThrGluAsp 340
Db 1185 CTGAGCAG 1244
Qy 341 GlySerIlybIlybPheValArgSerProIlybAlaGlnIlybIlybSerValGlyValIlyb 360
Db 1245 GGGAGCAG 1304
Qy 361 ValAlaArgGlyAlaArgAsnIlybProLeuGlnIlybProAlaIlybProAlaValIlyb 380
Db 1305 GTGGCCAG 1364
Qy 381 SerAlaIlybIlybAlaIlybSerProAlaLeuLeuProSerLeuAlaProLeuGlyIlyb 400
Db 1365 TCGCCCACTGCTGAG 1424
Qy 401 IlybIlybIlybIlybAlaIlybGlnSerGlnGlnIlybCysArgGlyLeuGlnAlaProLeu 420
Db 1425 AAGCAG 1484
Qy 421 AlaSerGlyIlybProValLeuGlyArgGluIlybLeuArgIlybLeuProAlaLeuVal 440
Db 1485 GCTTCAG 1544
Qy 441 TrpGlnSerGlnAlaGlnIlybProValGlnArgIlybGluIlybIlybIlybIlyb 460
Db 1545 TATAG 1604
Qy 461 HisIlybIlybTrpIlybIlybPheTrpGlnThr 470
Db 1605 CACCACCATTCACCACTTCTACAGAGCA 1634
RESULT 5
AR405917 1438 bp DNA linear PAT 18-DEC-2003
LOCUS AR405917
DEFINITION Sequence 5 from patent US 6630321.
ACCESSION AR405917
VERSION AR405917.1 GI:40154965
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.

REFERENCE 1 (bases 1 to 1438)
AUTHORS Scott, M., Zeng, W. and Wharton, K.
TITLE Naked cuticle genes and their uses
JOURNAL Patent: US 6630323-A 5 07-OCT-2003;
FEATURES
source location/Qualifiers
1. 1438
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 6,17e-112 Length: 1438
Score: 2493.00 Matches: 469
Percent Similarity: 99.79% Conservative: 0
Best Local Similarity: 99.79% Mismatches: 1
Query Match: 99.76% Indels: 0
DB: Gaps: 0
US-09-993-966-7 (1-470) x AR405917 (1-1438)
Qy 1 MetGlyIlybLeuIlybSerIlybProAlaAlaValCylbArgArgIlybSerProGlnIlyb 20
Db 8 ATGGGAACTTCACTTCAAGCGCGCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 67
Qy 21 AspSerPheAlaValSerAlaAlaTrpAlaArgIlybGlyIlybGlnIlybTrpIlybArg 40
Db 68 GACAGCTTCCGCTGAG 127
Qy 41 GlnArgCysProGlyIlybValSerGlyProArgGlnLeuArgLeuAlaGlyThrIlyb 60
Db 128 CAGGCTGCGCGGCGGTCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 187
Qy 61 ArgSerThrArgGluLeuValGlyAspValLeuAspThrLeuSerGlnIlybGluGlu 80
Db 188 GAAAGCAG 247
Qy 81 AspAspPheArgLeuGlnValAlaLeuProProGlnIlybTrpAspGlyLeuGlySerGly 100
Db 248 GACAGCTTGGCTGAG 307
Qy 101 AspGluIlybIlybMetGluArgValSerGlnProCysProGlySerIlybIlybGlnIlyb 120
Db 308 GATGAG 367
Qy 121 PheGlnIlybGlnIlybCysAspValSerMetGluGluAspSerArgGlnIlybTrpThrPhe 140
Db 368 TTGAGAGAGCTCAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 427
Qy 141 ThrLeuTrpAspPheAspAsnGlnIlybValThrArgGluAspIlybThrSerLeuLeu 160
Db 428 ACCCTGTATGACTTTGACCAACAGCAGAGTCACTCCGAGAGAGAGAGAGAGAGAGAG 487
Qy 161 HisThrIlybTrpGluValValAspSerSerValAsnHisSerProThrSerSerIlybMet 180
Db 488 CACACCATCTATGAG 547
Qy 181 LeuArgValIlybLeuThrValAlaProAspGlySerGlnSerIlybAspSerValLeuVal 200
Db 548 CTGGGGTAAAGCTCACCGTGGCCCCCGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 607
Qy 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIlybProThrGluAspLeu 220
Db 608 AATCAGGCTACCTGACAGAGCCGAGGAGCCGAGCAGAGAGAGAGAGAGAGAGAGAGAG 667
Qy 221 ArgSerTrpGluIlybIlybGlnArgAlaProLeuArgPheGlnIlybSerArgLeuGlu 240
Db 668 CGGAGCTGGAG 727
Qy 241 GlnSerGlyCysTrpThrIlybIlybValAspGluAsnIlybGluArgAsnIlybTrpLeu 260
Db 728 CAGTCTGCTGTACCACTTGGTATGAGAACATCGAGAGAGAGAGAGAGAGAGAGAGAG 787
Qy 261 AspLeuAlaGlyIlybGluAsnTrpThrSerGlnPheGlyProGlySerProSerValAla 280

Db 788 GATTCGCCGGGATGAACTACACGTCCTCCCAATTTGGGCTGGCTCCCTCCGCTGGCC 847

Qy 281 GlnlySerGlnleuProProArgThrSerAspProThrArgSerArgSerHisGluPro 300

Db 848 CAGAGCTGAGAACTGCCCCCGGACCTTCAATCCCACTCGATCTCGCTCCCATGAGCGG 907

Qy 301 GlnAlaIleHisIleProHisArgIleProGlnIleValAspProAlaSerPheHisPhe 320

Db 908 GAAGCCATCCACATCCACACCGAAACCCCAAGCCGTGAGCCGGCTCCCTCCACTTC 967

Qy 321 LeuAspThrProIleAlaIleValSerGlnleuGlnIleArgIleuArgIleThrGlnAsp 340

Db 968 CTGGACACCCCAATCGCCCAAGCTTCAAGAGCTTCAAGACGGCTCCGGGACACTCAGGAC 1027

Qy 341 GlySerIleHisIlePheValArgSerProIleValGlnIleValSerValGlyValGlyHis 360

Db 1028 GGGAGCAAGCACTTGTGAGGTCCCAAGGCCCAAGGCAAGAGTGTGGGTGGGCGAC 1087

Qy 361 ValAlaArgGlyValAlaGlnleuValProProleuGlyProAlaIleProAlaValSerPro 380

Db 1088 GTGGCCGAGAGGGGCAAGAAACAGCCCTCTGGAGCCCGCATCCCTGGGTGTCCGCC 1147

Qy 381 SerAlaHisIleuAlaAlaSerProAlaIleuLeuProSerIleuAlaProleuGlyHisIle 400

Db 1148 TCGGCCCACTGGCTGGCAGCCCGCCCTCTCTCCCTAGCCCCCTCGGACAG 1207

Qy 401 LysHisIleHisIleArgAlaIleValSerGlnIleGlnIleCysArgGlyIleuGlnAlaProleu 420

Db 1208 AAGCACAAGACCGAGCCAGAGAGCCAGAGGAGCTGCGGAGCTGAGGACCACTG 1267

Qy 421 AlaSerGlyIleProValIleuGlyArgGlnHisIleuArgIleuLeuProAlaIleuVal 440

Db 1268 GCTCTGAGTGGCTGCTGCTGGGAGGAGCAGCTGCGGAGCTGCGGCTGTGTGTG 1327

Qy 441 TyrGlnSerGlnAlaGlyIleProValGlnArgHisIleuHisIleHisIleGlnHis 460

Db 1328 TATGAGAGCCAGGCGGCGGAGCGGCTCAAGACATGAGCACCACCATGAACT 1387

Qy 461 HisHisIleThrHisIlePheTyrGlnThr 470

Db 1388 CACCACATTAACCACTTCTTACAGACA 1417

RESULT 6
AF343352 1416 bp mRNA linear ROD 29-MAR-2001

LOCUS AF343352 Mus musculus Nkd mRNA, complete cds.
AF343352
AF343352.1 GI:13487304

DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
MUS musculus (house mouse)
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1416)
Yan, D., Wallingford, J.B., Sun, T.O., Nelson, A.M., Sakanaka, C.,
Reinhard, C., Harland, R.M., Pantl, W.J. and Williams, L.T.
Cell autonomous regulation of multiple Dishevelled-dependent
pathways by mammalian Nkd
Proc. Natl. Acad. Sci. U.S.A. 98 (7), 3802-3807 (2001)
21173639
11274398
2 (bases 1 to 1416)
Yan, D., Wallingford, J.B., Sun, T.O., Nelson, A.M., Sakanaka, C.,
Reinhard, C., Harland, R.M., Pantl, W.J. and Williams, L.T.
Direct Submision
Submitted (26-JAN-2001) Research, Chiron Corporation, 4560 Horton
Street, Emeryville, CA 94608, USA
Location/Qualifiers
1. 1416
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"

CDS

1. 1416
/note="mMkd; similar to Drosophila naked cuticle"
/codon_start=1
/product="Nkd"
/protein_id="AAK27485.1"
/db_xref="GI:13487305"
/translation="MGKLSKPAUVCKRSPESGDSRAVAAARKGIEIWIGRORP
GSVSPQRLACTVGKTRRELVDVSVESDSQEWFTLYDPMNKREDITSL
KRMERUSEPQASKQLEFEELQDVSVESDSQEWFTLYDPMNKREDITSL
HTIYEVDSVSNHPTSSKTLRYLTVAPGSGSKSVLPNHDLOSTRPADTPAE
ELRGMEKKORAPLRFQDSHLEQDCHNCVDENIERNNHLDLAGIENYSCGPS
PSVAOSELEPRLSNFTRSRHPEAHTIPRRPQVDPGSPHLDTPPKASPLOR
LRGTQDSKHPVSPKAGKNNMGMCARARAKRPPLPTPTTHVSPAHLATSPALLP
TLAPLHKHKHRAKESQASCRGLQGLPAAAGSITWGREQVRELPAVVESQAGAV
ORHHHHHHHHHHHFFVOP"

ORIGIN

Alignment Scores:

Pred. No.:	1,95e-97	Length:	1416
Score:	2192.50	Matches:	410
Percent Similarity:	91.06%	Conservative:	18
Best Local Similarity:	87.23%	Mismatches:	41
Query Match:	87.74%	Indels:	1
DB:	10	Gaps:	1

US-09-993-966-7 (1-470) x AF343352 (1-1416)

Qy 1 MetGlyLysLeuHisSerIleValAlaValCysLysArgArgIleuSerProGlnGly 20

Db 1 ATGGGGAATCTTCACTGAAAGCGGCGCGCTGTGCAAGGCGAGGAGCCCGGAAGGT 60

Qy 21 AspSerPheAlaValSerAlaAlaTPAlaArgLysGlyIleGlnGluTTP1IleGlyArg 40

Db 61 GACAGCTTCTCTGAACCGCTGTGGGCAAGAAAGCATCGAGGTGATCGGAGG 120

Qy 41 GlnArgCysProGlyIleValSerGlyProArgGlnleuArgIleuArgIleGlnIleGly 60

Db 121 CAGGCTGTCCAGGACCGCTCTCAAGACCCGCTCACTGATGATTGGACGACTGTGGT 180

Qy 61 ArgSerThrArgGlnleuValGlyAspValLeuArgAspThrIleuSerGlnGlnGlu 80

Db 181 CGAGGCACTCGGAACCTCGTGGGTGACACTTCAAGAGGCTCTCGGTGAGGAGAGAG 240

Qy 81 AspAspPheArgIleuGlnValAlaIleuProProGlnIleThrAspGlyIleuGlySerGly 100

Db 241 GACGACTTCCCTTAGAAGTGGCCCTGCGCTGAGAAAGATCAACAGCTTAGGTAGTGA 300

Qy 101 AspGlnLysIleMetGlnArgValSerGlnProCysPheProIleSerLysGlnIleuLys 120

Db 301 GATGAGAGAGAAATGGAGAGACTGAGGAACTGGCCAGGCTTCAAGAGCAGCTCAG 360

Qy 121 PheGlnGlnleuGlnCysAspValSerMetGlnLysAspSerArgIleuGlnIleuThrPhe 140

Db 361 TTGGAAGAGTACAGTGTATGTCTGTGTGAGAGAGACAGCCGGCAAGAGTGAATTC 420

Qy 141 ThrLeuTyrAspPheAspAsnAngIleValThrArgGlnAspIleThrSerLeuLeu 160

Db 421 ACTCTATATGACTTCGCAACAATGGCAAGTACCGGTAGGACATTAACAGCTTCTG 480

Qy 161 HisThrIleTyrGlnValValAspSerSerValAspHisSerProThrIleSerIleMet 180

Db 481 CATACCATCTATAGAAGTGTGACTCTGTGTAAACCATTCCTCCCATCAAGCAAGACA 540

Qy 181 LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysArgSerValLeuVal 200

Db 541 CTGCGGTGAAGCTCACCGTGTCTCTGAGAGGAGCAAGATGAAGGAGCGTCTTTC 600

Qy 201 ArgGlnAlaAspIleuGlnSerAlaArgProArgAlaGlnIleThrLysProThrGlnAspLeu 220

Db 601 AACCATACCATTCGAGAGCAAGAGCCCGGAGCAGACCAAAACCCGCTGAGAGCTG 660

Qy 221 ArgSerTyrGlnLysIleValArgAlaProleuArgPheGlnGlnIleAspSerArgLeuGlu 240

```

Db      CGTGGCTGGAGAGAGACGACGACCCCACTCAGTTCCAGGGTGAACGACCTGGAG 720
Qy      GlnSerGlyCysTyrHisGlyValAspGlnAsnIleGluArgArgAsnHisTyrLeu 260
Db      GlnSerGlyCysTyrHisGlyValAspGlnAsnIleGluArgArgAsnHisTyrLeu 260
Db      CACCCAGACTGCTACACCCCATTTGGTGGATGAGAACATTGAGAGGAAACCACTACCTA 780
Qy      AspleuAlaGlyIleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValAla 280
Db      GACTGGCGGGGATGAGAACCTACACGCTCTGATTGACCGGGATCTCCCTCGGTGGCC 840
Qy      GlnLysSerGlnLeuProProArgThrSerAsnProThrArgSerArgSerHisGluPro 300
Db      CAGAGTCAAGTAGCTGGCCCTCGAATCTCCAAACCCCACTGGCTCTCGCTCCCAAGACCA 900
Qy      GluAlaIleHisIleProHisArgGlyProGlnGlyValAspProAlaSerPheHisPhe 320
Db      GAAGCTGCCCACTCCACACACCGGAGGCCCAAGGTGAGACCAAGGCTCTTCACACTC 960
Qy      LeuAspThrProIleAlaLysValSerGlnLeuGlnGlnArgGlyThrGlnAsp 340
Db      CTGGACACCCCATTTGGCCAGAGGCAATCAGAGCTCCAGACAGCGCTCCGGGCACTCAGGAT 1020
Qy      GlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValGlyHis 360
Db      GGGAGCAGACACTTGTGAGGTCTCCCAAGGCCAGGGCAAGAACATGGGTATGGGCAC 1080
Qy      ValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
Db      GGGGCGCCAGAGGTGCAAGACAGACAGCCCTGCTACCCACCACTACTGCTCTCCCCC 1140
Qy      SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHis 400
Db      TCTGCCATCTGGCCACACAGCCCTCTCTCCCACTGSCACCCCTCGGGCACAAG 1200
Qy      LysHisLysHisArgAlaLysGlySerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
Db      AAACACACAGCATCGAGCCAGAGACAGCCAGGCTGCGGGGCTCGACAGGCCCCCTG 1260
Qy      AlaSerGlyGly--ProValLeuGlyValArgGlnHisLeuArgGlyLeuProAlaLeuVal 439
Db      GCGCAGAGAGGCTCCACCGTATGGGCGGAGAGCAGAGGAGCTGCGCTGGTGGTG 1320
Qy      ValTyrGlnSerGlnAlaGlyGlnProValGlnArgHisGlnHisHisGlnHisGln 459
Db      GTGTACAGAGCCAGGCTGGGCGAGCCGCTCCAGAGACCAACCACTCACACACGAA 1380
Qy      HisHisHisHisTyrHisHisPheTyrGln 469
Db      CATCACACCATTTATCACCACTTCTATCAG 1410

```

RESULT 7
BC034838
LOCUS
DEFINITION
 Mus musculus naked cuticle 1 homolog (Drosophila), mRNA (cDNA clone MGC:40995 IMAGE:1429943), complete cds.
ACCESSION
BC034838
VERSION
BC034838.1 GI:22028144
KEYWORDS
SOURCE
 Mus musculus (house mouse)
ORGANISM
 Mus musculus
REFERENCE
AUTHORS
 Strauberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buettow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Udell, T.B., Toshitsuki, S., Carninci, P., Prange, C., Raha, S.S., Loguettano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,

TITLE
 JOURNAL
 PUBMED
 REFERENCE
 AUTHORS
 JOURNAL
 DIRECT Submission
 Submitted (29-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK
 COMMENT
 Email: egads-remail.nih.gov
 Tissue Procurement: Marcello Bento Soares, Ph.D.
 cDNA Library Preparation: M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
 Info@cgsc.bc.ca
 Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield, Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim Macdonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin, Teika Olson, Diana Palmquist, Anca Petrescu, Anna Lisa Prehbu, Parvaneh Saeedi, Ur Santos, Angeline Scherch, Ursula Skalske, Duane Smaluis, Jeff Scott, Miranda Tsai, George Yang, Jacqui Schein, Asim Siddiqui, Rob Holt, Marco Marra.
FEATURES
 source
 1..1765
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="MGC:40995 IMAGE:1429943"
 /tissue_type="Mammary gland"
 /clone_lib="Soares_mammary_gland_NbMWG"
 /lab_host="DH10B"
 /note="Vector: pT73-Pac"
 1..1765
 /gene="Nkd1"
 /db_xref="locusid:93960"
 /db_xref="mgi:2135954"
 138..1353
 /gene="Nkd1"
 /codon_start=1
 /product="naked cuticle 1 homolog"
 /protein_id="AAH34838.1"
 /db_xref="gi:22028145"
 /db_xref="locusid:93960"
 /db_xref="mgi:2135954"
 /translacion="MGKLSKPAVCKRRSPGDSFVSAAMARKGIEBWMGRQPC
 KNSGPROLACTGVGRTRELQDVSRELAEGEDEDPLEVALPEKIDLSGDB
 GMEKRLSEPGQAKKOLKFEELQDVSVESDSQEMFTLYDPNNKXVRBDTSL
 HTIYVDSVSNVSPSSKTLRYKLTVAAPDSQSKSLVNNHNDIOSRPRADKPAE
 ELWGSEKQKQAPRFQDSHLEQPDCTHCVENIERNNHYLDLAGIENTSOGPSS
 PSVAQKSELPKPSINPFRSHSEPEAAHVHRPPQGVDPSEFILDTPFAVASLQOR
 LRDTQDSKIFVSPRAQGNMGMGARBARSPPLVPTTHVTSFAHLATSPALP
 TLAFLGHKKHKBKSKSQASCRGIGPLAAGSGTVMGEQVRELPVAVVYESQAGQAV

ORHHHHHHHHHHHHFYOP"

Alignment Scores:

Pred. No.: 2,81e-97 Length: 1765
Score: 2191.50 Matches: 409
Percent Similarity: 91.06% Conservative: 19
Best Local Similarity: 87.02% Mismatches: 41
Query Match: 87.70% Indels: 1
DB: 10 Gaps: 1

US-09-993-966-7 (1-470) x BC034838 (1-1765)

QY 1 MetGlyLeuHisSerLyProAlaAlaValCyLeuSerArgGlyUserProGluGly 20
Db 138 ATGGGAAATCTCACTGAAGCCGGCCGCTGTGCAAGCGCAGAGAGCCCGGAAAGGT 197
QY 21 AspSerPheAlaValSerAlaAlaTTPAlaArgLyGly1LegLutTP1LegLYArg 40
Db 198 GACAGCTTCTGTAAACGCTGCTTGGGCAAGAAAGCATCGAGAGTGGATGGAGG 257
QY 41 GlnArgCyProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThr1LegLy 60
Db 258 CACGCTGTCCAGGCAAGCGCTCTCAGACCCGCTCAGCTGAGATTGGCAGGCACTTGGT 317
QY 61 ArgSerThrArgGlnLeuValGlyValPheValLeuArgPheThrLeuSerGlyGluGlu 80
Db 318 CGAGGCACTGGGAACTGTGGGTGACCTTGAAGAGGCTCTCGGTGAGAGAGAGAG 377
QY 81 AspAspPheArgLeuGlnValAlaLeuProProGlyUtrThrArgGlyLeuGlySerGly 100
Db 378 GAGGACTTCCCTAGAAAGTGGCCCTCCGCTGAAAGATGACACCTAGGTAAGTGA 437
QY 101 AspGlyUtrLeuSerGlyUtrValSerGlyProCySerProGlySerLyGlnLeuLy 120
Db 438 GATGAGAGAGAGATGAGAGACTGAGGAGCACTGCGCAGGCTCCAGAAAGCAGCTCAG 497
QY 121 PheGlnGluLeuGlnCyAspValSerMetGluGlnLeuSerArgGlnUtrTPhe 140
Db 498 TTTGAAAGAGTGAAGTGTGATGCTCTGTGAGAGAGACAGCCGCAAGAGTGAAGTTC 557
QY 141 ThrLeuTyAspPheAspAsnAngLyLeuValThrArgGlnAsp1LeuThrSerLeu 160
Db 558 ACTCTATGACTCTTCAACAACATGCAAGTGAACCTGAGACATTAACAGCTTCCTG 617
QY 161 HisThr1LeuGlyValValAspSerSerValAsnHisSerProThrSerSerLyMet 180
Db 618 CATACCATCTATGAAAGTGTGACTCTCTGTGAACATTCCCCCATCAAGCAAGACA 677
QY 181 LeuArgValLyLeuThrValAlaProArgGlySerGlnSerLyBAGSerValLeuVal 200
Db 678 CTGCGGGTGAAGTCTACCGGTGCTCTGAGAGGAGCAAGTGAAGAGACCGCTTTTC 737
QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrArgProThrArgLeu 220
Db 738 AACCAATCCGATCTGCAAGACACAGGCCCGCAGACACACCAACCCGCTGAGAGCTG 797
QY 221 ArgSerTrpGlyUtrLySerGlnArgAlaProLeuArgPheGlnLyAspSerArgLeuGlu 240
Db 798 CGTGGCTGGGAGAAAGAGAGAGAGGCCCACTCAGAGTTGCAAGGTGACACCACTGGAG 857
QY 241 GlnSerGlyCySerGlyHisGlyValAspGlnLeuArgGlnArgAsnHisGlyLeu 260
Db 858 CAGCCAGACTGCTACCAACCACTTCGTGAGAGAACTTGAAGAGAGAAACCACTACTA 917
QY 261 AspLeuAlaGly1LegLutTP1SerGlnPheGlyProGlySerProSerValAla 280
Db 918 GACCTGGCGGAGTGAAGATCAAGCTCTCAAGTTGACCGGAGTCCCTTCGGTGGCC 977
QY 281 GlnLySerGlyUtrLeuProArgTrpThrSerAsnProThrArgSerArgSerHisGluPro 300
Db 978 CAGAAATGAGAGCTGCGCCCTCGAATCTCAACCCCACTGCTCTGCTCCCAAGAGCCA 1037

QY 301 GlnAla1LeuHis1LeuProHisArgLySerProGlnGlyValAspProAlaSerPheHisPhe 320
Db 1038 GAACTGCCACAGCTGCCACACCGAGGCCCGCAAGTGTGACCCAGGCTCTTCCACCTC 1097
QY 321 LeuAspThrPro1LeuAlaLyValSerGlyLeuGlnGlnArgLeuArgLyThrGlnAsp 340
Db 1098 CTGACACCCCATTTGGCAAGGATCAGAGCTCCAGACACCGCTCCGGGAGCACTAGGAT 1157
QY 341 GlySerLyHisPheValAspSerProLyAlaGlnLyLySerValGlyValGlyHis 360
Db 1158 GGGAGCAAGCACTTTGAGAGTCCCGCAAGGCCCGCAAGAAACATGAGTATGGGCAC 1217
QY 361 ValAlaArgGlyAlaArgAsnLyProProLeuGlyProAla1LeuProAlaValSerPro 380
Db 1218 GGGGCCAGAGTGCMAAGAACAGCTTCACTGTATCCACACCACTACTGTCTCCCC 1277
QY 381 SerAlaHisLeuAlaAspSerProAlaLeuLeuProSerLeuAlaProLeuGlyHis 400
Db 1278 TGTGCCATCTGGGCACACAGCCCAAGCCCTTCTCCCACTGGCAGCCCTGGGCAAG 1337
QY 401 LyHisLyHis1ArgAlaLyGlySerGlnGlnLyCyAspArgLyLeuGlnAlaProLeu 420
Db 1338 AAACACAAAGCATGAGCCCAAGAGAGCCAGGCTGCGGAGGCTGAGGCGCCCTG 1397
QY 421 AlaSerGlyGly---ProValLeuGlyArgGlnHis1LeuArgGlyUtrLeuProAlaLeuVal 439
Db 1398 GCTGCAAGAGCTTCCACCTGATGGGGGAGACAGGTGAGGAGAGCTGCTGCGGTG 1457
QY 440 ValTrpGlnSerGlnAlaGlyGlnProValGlnArgHisGlnHis1HisGln 459
Db 1458 GTGTACAGAGCCAGAGTGGGCAAGCCCTCCAGAGACACAAACCACTACACCAAGAA 1517
QY 460 HisHisHis1GlyHis1HisPheTyGln 469
Db 1518 CATCACCAACCATTAACCACTTCTATCAG 1547

RESULT 8

AF358134 1671 bp mRNA linear ROD 29-MAY-2001
LOCUS AF358134
DEFINITION Mus musculus naked cuticle-1 (Nkd1) mRNA, complete cds.
ACCESSION AF358134
VERSION AF358134.1 GI:14211711
KEYWORDS

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 1671)
Wharton, K.A. Jr., Zimmermann, G., Rouselet, R. and Scott, M.P.
Vertebrate proteins related to Drosophila Naked Cuticle bind
dishevelled and antagonize Wnt signaling
Dev. Biol. 234 (1), 93-106 (2001)

JOURNAL

2 (bases 1 to 1671)
Wharton, K.A. Jr., Zimmermann, G. and Scott, M.P.
Direct Submision
Submitted (08-MAR-2001) Pathology and Molecular Biology, University
of Texas Southwestern Medical School, 5323 Harry Hines Blvd.,
Dallas, TX 75390-9072, USA

FEATURES

Location/Qualifiers
1..1671
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6XCHRA"
/db_xref="taxon:10090"
/clone="L84"
/sex="female"
/tissue_type="lung"
/dev_stage="6-8 weeks"
1..1671
/gene="Nkd1"
80..1495

source

1..1671
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6XCHRA"
/db_xref="taxon:10090"
/clone="L84"
/sex="female"
/tissue_type="lung"
/dev_stage="6-8 weeks"
1..1671
/gene="Nkd1"
80..1495

gene

1..1671
/gene="Nkd1"
80..1495

CDS

80..1495


```

/gene="Nkd1"
/note="contains EFX domain; similar to Drosophila
melanogaster naked cuticle (nkd)"
/codon_start=1
/product="naked cuticle-1"
/protein_id="AKS7483.1"
/db_xref="GI:14211712"
/translation="MGKIKSPAAVCKRPSPGSDSPAVSAAMKGIIEIGRORCP
GSVSGRQRLACTVGRGTRRELYDTSREALGEDEDDPFLRYALPREKIDISGSDP
KRMERLSEPEQASKKQKFEELQDVSEDSQKSVLFNHDQSTRPADTSPLE
HTLYEVDSVSNHPTSSKTLRYALVADPGSQSKSVLFNHDQSTRPADTSPLE
ELRGWKQKAPLRFOGDSHLEQPDCHNCVDENIRNHYLDLGIENTSGPGPS
PSVAQSELPRIISNPTSRSHPEALHPIRRPGVDPSFLIDTPFAKSELQOR
LRGTQDSKHFRVSPKAGKMGHGARGARASKPLVPTHTVSPSHLATSPLLP
TLAPLGHKHKHAKESQASCRGIQPLAAGSTWGRREVRLPAMVYESQAGAV
QRHHHHHHHHHHHFTPT"

```

ORIGIN

Alignment Scores:

Pred. No.:	3.3e-97	Length:	1671
Score:	2189.50	Matches:	409
Percent Similarity:	91.06%	Conservative:	19
Best Local Similarity:	87.02%	Mismatches:	41
Query Match:	87.62%	Indels:	1
DB:	10	Gaps:	1

US-09-993-966-7 (1-470) x AF358134 (1-1671)

```

QY 1 MetGlyLysLeuHisSerIysProAlaAlaValCysIysArgArgIuseProGluGly 20
DB 80 ATGGGGAACTTCCTCGAAGCCGGCCGCTGTGCAAGCCAGAGAGAGCCCGAAGGT 139
QY 21 ArgSerPheAlaValSerAlaAlaTTPAlaArgIysGlyIleGluGluTTPAlaArg 40
DB 140 GAAGAGCTTTCCTGTAACCGCTGCTTGGGCAAGAAAGCATCGAGAGTGGATCGGAGG 199
QY 41 GluArgCysProGlyGlyValSerGlyProArgGluLeuArgLeuAlaGlyThrIleGly 60
DB 200 CAGCGCTGTCCAGGACCGCTCTCAGAACCCCGTCACTGAGATTGGCAGGACCTTGGT 259
QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgPheThrIleuSerGluGluGlu 80
DB 260 CGAGGACCTCGGAACTCGTGGGTGACACTTCAGAGAGAGGCTTCGGTGAAGAGAGAG 319
QY 81 AspArgPheArgLeuGluValAlaLeuProProGluIysThrArgGlyLeuGlySerGly 100
DB 320 GACGACTTCCCTCAGAGAGTGGCCCTGCGCTGAGAGAGATCGACACCTTAGGTAGTGA 379
QY 101 AspGluLysLysMetGluArgValSerGluProCysProGlySerIysLysGluLeuLys 120
DB 380 GATGAGAGAGATGAGAGACTGAGCGAACCTGGCCAGGCTCCAGAGACCTCAG 439
QY 121 PheGluGluLeuGluCysAspValSerMetGluGluAspSerArgGluIysThrPhe 140
DB 440 TTGGAAGAGTCAAGTGTGATGTCTGTGAGAGAGACAGCCGCAAGAGTGAAGTTC 499
QY 141 ThrLeuIysArgPheAspPheAspArgGlyLysValThrArgGluAspIleThrSerLeuLeu 160
DB 500 ACTCTAATAGACTTCGACCAACAAAGCAAGTGAACCGTGAAGCAATTACCAAGCTTCG 559
QY 161 HisThrIleArgValValAspSerSerValAsnHisSerProThrSerSerLeuMet 180
DB 560 CATTCATCTATGAGAGTGTGACTCTCTGTGAAACATTCCTCCCACTCAAGAGAGACA 619
QY 181 LeuArgValIysLeuThrValAlaProAspGlySerGluSerIysArgSerValLeuVal 200
DB 620 CTGGGGGTGAAGCTCACCGTGGCTCTGACGGGAGCCAGAGTAAAGAGAGGCTCTTTC 679
QY 201 AsnGluAlaAspLeuGluSerAlaArgProArgAlaGluThrIysProThrGluAspLeu 220
DB 680 AACCATACCATCTGACAGACCAAGGCCCGGAGCAGACACCAAAACCGCTGAGAGAGCTG 739
QY 221 ArgSerTrpGluLysLysGluArgAlaProLeuArgPheGluGluIysAspSerArgLeuGlu 240

```

```

DB 740 CCGGGCTGGAGAAAGAGCAGAGCCCACTCAGGTTCCAGGGTGCAGACCTGGAG 759
QY 241 GlnSerGlyCysArgThrHisSerCysValaArgGluAsnIleGluArgArgAsnHisIleLeu 260
DB 800 CACCCAGACTGCTACACCACTTCGCGGAGTGAAGAACATGAGAGAGAAACCACTACTTA 859
QY 261 AspLeuAlaGlyIleGluAsnIysThrSerGluPheGlyProGlySerProSerValAla 280
DB 860 GACTGGCGGGATAGAGAACTACACTCTCACTTTGGACCGGGATCCCTTGGTGGCC 919
QY 281 GlnLysSerGluLeuProProAlaGluThrSerAspProThrArgSerArgSerHisGluPro 300
DB 920 CAGAAAGTCAGAGCTGCCCTCCCAATCTCAACCCCACTGCTGCTCCCAAGAGCCA 979
QY 301 GluAlaIleHisIleProHisArgIysProGluGluValaAspProAlaSerPheHisPhe 320
DB 980 GAAGCTGCCCACTCCACACCGGAGGCCCAAGGTGTGAGCCAGGCTCTTCCACTTC 1039
QY 321 LeuAspThrProIleAlaLysValSerGluLeuGluGluArgLeuArgGlyThrGluAsp 340
DB 1040 CTGACACCCCAATTTGCAAGGATCAGACTCAGACGCTCCGGGCACTCAGAGAT 1099
QY 341 GlySerLysHisArgPheValaArgSerProLysAlaGluGlyLysSerValGlyValHis 360
DB 1100 GGGAGCAGACACTTGTGAGGTCCCAAGGCCCCAAGGCAAGAACATGAGTATGGGCAC 1159
QY 361 ValAlaArgGlyValaArgAsnLysProProLeuGlyProAlaIleProAlaValaSerPro 380
DB 1160 GGGGCCAGAGTCCAGAAAGAACAGCCCTCCACTGTATCCCAACCCACTACTGTCTCCCC 1219
QY 381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
DB 1220 TCTGCCATTTGGCCACAGCCCAAGCCCTTCTCCCACTTGGCACTCCCGGCAAG 1279
QY 401 LysHisLysHisArgAlaLysGluSerGluGluGlyCysArgGlyLeuGluAlaProLeu 420
DB 1280 AAACACAGCATCGAGCCAGAGAGAGCAGGAGCTGCGGGGCTCGAGGGCCCCCTG 1339
QY 421 AlaSerGlyLys---ProValaLeuGlyArgGluHisIleuArgGluLeuProAlaLeuVal 439
DB 1340 GCTGCAGAGAGCTCCACCTGATGGGGGGAGCCAGGTGAGGAGCTGCTGCTGATG 1399
QY 440 ValTyrGluSerGluAlaGlyGlnProValaGluArgHisGluHisHisIleGlu 459
DB 1400 GTGTACAGAGCCAGGCTGGGAGCGCTCAGAGACAGACACATCACACAGAA 1459
QY 460 HisHisHisIstYrHisHisPheTyrGln 469
DB 1460 CATCACCATATATCACCACTTCTATCAG 1489

```

RESULT 9

AR405916

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. (bases 1 to 1731)

Scott, M., Zeng, W. and Wharton, K.

Naked cuticle genes and their uses

Patent: US 6630323-A 3 07-OCT-2003;

Location/Qualifiers

1..1731

/organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Alignment Scores:

Pred. No.:	1.82e-96	Length:	1731
------------	----------	---------	------

Score: 2174.50 Matches: 407
Percent Similarity: 90.64% Conservative: 19
Best Local Similarity: 86.60% Mismatches: 43
Query Match: 87.01% Indels: 1
DB: 6 Gaps: 1

US-09-993-966-7 (1-470) x AR405916 (1-1731)

QY 1 MercGlyLeuLeuHisSerLySPRoAlaValCylValSerArgArgGluSerProGluGly 20
Db 140 ATGGGGAAACCTTCACTCGAAGCCGGCCGGTGTGCAAGCGCAGAGGAGAGCCCGAAGGT 199
QY 21 AspSerPheAlaValSerAlaAlaTPAlaArgGlyGlyLeuGluTTPLeuGlyArg 40
Db 200 GACAGCTTGTCTGTAGCGCTGCTGGCAAGAAAGCATCGAGAGTGGATCGGAGG 259
QY 41 GlnArgCySPProGlyValSerGlyProArgGluLeuArgLeuAlaGlyThrTlGly 60
Db 260 CAGCGCTGTCAAGGACGCTCTCAAGACCCCGTCACTGAGATTGGCAGGACCTGTGGT 319
QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80
Db 320 CGAGGCACTCGGAAACTCGTGGGTGACACTTCTAGAGAGGCTCTCGGTGAGAGAGAG 379
QY 81 AspAspPheArgLeuGluValAlaLeuProProGluLeuThrAspGlyLeuGlySerGly 100
Db 380 GACGACTTCCCTTGAAGAGTGGCCCTGCGCTGAGAGAGATCGACAGCTAGGTAGTGA 439
QY 101 AspGluLeuLeuSerGluArgValSerGluProCySPProGlySerLySGLLeuLys 120
Db 440 GATGAGAGAGAAATGAGAGAGACTGAGGAACTGGCCAGGCTCTCAAGAAAGCAGCTCAG 499
QY 121 PheGluGluLeuGlnCySPAspValSerMetGluGluAspSerArgGlnGluTTPThrPhe 140
Db 500 TTGGAAGAGCTACAGTGTGATGTCTGTGAGAGAGACAGCCGCAAGAGTGGACTTTC 559
QY 141 ThrLeuTyRAspPheAspAsnGlyLysValThrArgGluLeuPLeuThrSerLeuLeu 160
Db 560 ACTCTATATGACTTCCAAACAATGAGCAAGTACCCGTAGAGACATTACAGCTGCTGCTG 619
QY 161 HisThrTleryGluValValAspSerSerValAsnHisSerProThrSerSerLyMet 180
Db 620 CATACCATCATTAAGAGTGTGACTCTCTGTGAACATTCCCCCATCATACAGCAGACA 679
QY 181 LeuArgValLeuThrValAlaProAspGlySerGlnSerLySGLLeuVal 200
Db 680 CTGCGGTGAACTCAGCCGTGCTCTGACGAGGACAGATGAAGAGACGCTCTTTC 729
QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLySPProThrGluAspLeu 220
Db 740 AACCATACCATTTGACAGACACAGGCCCGGAGCAGACCAAAACCGCTGAGAGCTG 799
QY 221 ArgSerTTPGluLeuValSerGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
Db 800 CCGGCTGGGAGAAAGCAGACGCCCTCAGGTTCCAGGTGTACAGCCACTGAG 859
QY 241 GlnSerGlyCySPThrHisHisCySPValAspGluAsnLeuGluArgArgAsnHisGlyLeu 260
Db 860 CAGCCAGACTGTACACCATCTGCGTGTGATGAGAACTTGAAGAGAAACCACTACTA 919
QY 261 AspLeuAlaGlyLysLeuLeuValThrSerGlnPheGlyProGlySerProSerValAla 280
Db 920 GACTGCGCGGAGTAAAGAACTACAGCTTCAGTTTGGACCGGAGATCCCTTGGGTGGCC 979
QY 281 GlnLySerGlyLeuLeuProProArgThrSerAsnProThrArgSerSerArgSerHisGluPro 300
Db 980 CAGAATTCAGAGTGGCCCTCTCAATCTCAACCCCACTGCTCTCCCTCCAGAGCCA 1039
QY 301 GlnAlaTlHisHisLeuProHisArgLySPProGlnGlyValAspProAlaSerPheHisPhe 320
Db 1040 GAAGCTGTGCTGTAAAGCGCTGTGGCAAGAAAGCATCGAGAGTGTGATCGGAGG 1099
QY 321 LeuAspThrProTlalaValSerGluLeuGlnGlnArgLeuArgGlyThrGlnAsp 340

Db 1100 CTTGACACCCCATTTGGCCAAAGCATTCAGACTTCCAGCAAGCGCTCCGGGACATCAGAT 1159
QY 341 GlySerLySHPheValArgSerProLyAlaGlnGlyLySerValGlyValGlyHis 360
Db 1160 GGGAGCAAGCACTTTGTGAAGTGTCCCAAGGCCAGGCAAGAAACATGGGTATGGCCAC 1219
QY 361 ValAlaArgGlyAlaArgAsnLySPProLeuGlyProAlaTlProAlaValSerPro 380
Db 1220 GGGGCCAGAGGTGCAGAAACCAAGCTTCAGTGTATCCACCAACCATTAAGTCTCCCC 1279
QY 381 SerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHis 400
Db 1280 TGTGCCATCTGGCCACAGCCAGCCCTTCTCCCACTGACCTTGGGACCCCTGGGACAG 1339
QY 401 LySHisLySHPheAlaValSerGlnGlnGlyCySPArgGlyLeuGlnAlaProLeu 420
Db 1340 AAACACAAAGCATGAGGCAAGAGAGCAGGCGAGCTGCGGGGCTGCAAGGCGCCCTG 1399
QY 421 AlaSerGlyLySPProValLeuGlyArgGluHisLeuArgGluLeuProAlaLeuVal 439
Db 1400 GCTGCAAGAGGCTTCCACCGTCAATGAGGCGGAGCAGGTGAGGGAGCTGCTGCGGTATG 1459
QY 440 ValTyRArgLeuGlnAlaGlyInProValGlnArgHisGluHisHisHisGlu 459
Db 1460 GTGTACAGAGCCAGAGCTAGGCGAGCGCTCCAGAGACAGAAACATCACACACAGAA 1519
QY 460 HisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 469
Db 1520 CATCACCATTTATCACCACTTCTATCAG 1549

RESULT 10
AX172351 1401 bp DNA linear PAT 03-JUN-2001
LOCUS AX172351
DEFINITION Sequence 1 from Patent WO0144279.
ACCESSION AX172351
VERSION AX172351.1 GI:14597503
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
1 Yan, D. and Williams, L. T.
Mammalian dishevelled-associated proteins
JOURNAL Patent: WO 0144279-A 1 21-JUN-2001;
Chiron Corporation (US)

FEATURES
source Location/Qualifiers
1..1401
/organism="Mus musculus"
/mol_type="unassigned DNA"
/db_xref="taxon:10090"

ORIGIN
Alignment Scores: 3.09e-96 Length: 1401
Pred. No.: 2167.50 Matches: 406
Score: 2167.50 Conservative: 18
Percent Similarity: 90.79% Mismatches: 42
Best Local Similarity: 86.94% Indels: 1
Query Match: 86.73% Gaps: 1

US-09-993-966-7 (1-470) x AX172351 (1-1401)

QY 1 MercGlyLeuLeuHisSerLySPRoAlaValCylValSerArgArgGluSerProGluGly 20
Db 1 ATGGGAAACCTTCACTCGAAGCCGGCCGGTGTGCAAGCGCAGAGGAGAGCCCGAAGGT 60
QY 21 AspSerPheAlaValSerAlaAlaTPAlaArgGlyGlyLeuGluTTPLeuGlyArg 40
Db 61 GACAGCTTGTCTGTAGCGCTGCTGGCAAGAAAGCATCGAGAGTGTGATCGGAGG 120
QY 41 GlnArgCySPProGlyValSerGlyProArgGluLeuArgLeuAlaGlyThrTlGly 60


```

QY 207 SerAlaArgProArgAlaGluThrLysProThrGluAspLeuArgSerTyrGluLys 226
Db 442 AGGCAAGAGGCGCCGAGAGAGACCAACCACTGAGGACCTGGGAGGAGAAAG 501
QY 227 GlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlnInsArgIleCysTyrHis 246
Db 502 CAGCGAGCGCCGCTCAGGTTCCAGGGTGCAGACGCCCTGAGAGAGCTGTGGCTGACAC 561
QY 247 HisCysValaAspGluAsnIleGluArgArgAsnIleTyrLeuAspLeuAlaGlyIleGlu 266
Db 562 CATTGGCTGATGAGAACTGAGAGAGAGAAACCACTTAGATCTCGCCGGGATGAA 621
QY 267 AsnTyrThrSerGlnPheGlyProGlySerProSerValaAlaGlnLysSerGlnLeuPro 286
Db 622 AACTACACGCTCCCAATTGGGCTGCTGCTCCCTTCCTGAGCCCAAGAGTGAACCTGCC 681
QY 287 ProArgThrSerAspProThrArgSerArgSerIleGluProGluAlaIleHisIlePro 306
Db 682 CCCCCGACCTCCCAATCCCACTGCATCTGCTCCCATAGCCGAGAGCCATCCATCCCA 741
QY 742 HisArgLysProGlnGlyValaAspProAlaSerPheHisPheLeuAspThrProIleAla 326
Db 327 LysValaSerGlnLeuGlnGlnArgLeuArgGlyThrGlnAspGlySerIleHisPheVal 346
Db 802 AAGGTCACAGAGCTCCAGCAACGCTCCGGGAGCCAGAGACGGAGCAACGACTTTGTG 861
QY 347 ArgSerProLysAlaGlnGlyLysSerValaGlyValaGlyHisValaAlaArgGlyAlaArg 366
Db 862 AGGTCCCGCCAGGCGCCAGGCGGAGAGTGTG----- 891
QY 367 AsnLysProLeuGlnProAlaIleProAlaValaSerProSerAlaHisLeuAlaAla 386
Db 891 ----- 891
QY 387 SerProAlaLeuLeuProSerLeuAlaProLeuGlyHisIleLysLysHisIleArgAla 406
Db 892 -----GCCCTCCGCGGACAGAGCAAGAGCAAGCAACCGAGCC 927
QY 407 LysGlnSerGlnGlnGlyCysArgGlyLeuGlnAlaAlaProLeuAlaSerGlyGlyProVal 426
Db 928 AAGGAGAGCCAGAGGGGCTCCGGGCTTCAGAGCACCACTGGCTCAGGTGGCCCTGTC 987
QY 427 LeuGlyArgGlyHisLeuArgGluLeuProAlaLeuValaTyrGlnSerGlnAlaGly 446
Db 988 CTGGGGCGGAGACCTGGGAGAGCTGCCGCTTGTTGTATGAGACCAAGCGCGG 1047
QY 447 GlnPro 448
Db 1048 CAGCCG 1053

```

RESULT 12
AC145175 207486 bp DNA linear HTG 15-AUG-2003
AC145175
DEFINITION Pan tlog1odytes clone CH251-32B1, WORKING DRAFT SEQUENCE, 11
ordered pieces.
AC145175
VERSION AC145175.2 GI:33667124
KEYWORDS HTG, HTGS_PHASE2, HTGS_DRAFT.
SOURCE Pan tlog1odytes (chimpanzee)
ORGANISM Pan tlog1odytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
1 (bases 1 to 207486)
Antoniellis, A., Ayele, K., Beckstrom-Stenberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Carriaga, K.,
Chu, G., Coleman, B., Coleman, H., Engle, J., Granite, S., Guan, X.,
Gupte, J., Haghighi, P., Han, J., Hansen, N., Ho, S.-D., Hu, P.,
Hurle, B., Idol, J.R., Karlins, E., Kwong, P., Latic, P., Lee-Lin, S.-O.,
Legardi, R., Maduro, Q.L., Maduro, V.B., Margulies, E.H., Massello, C.,
Maskeri, B., McDowell, J., Paguitigan, C., Pearson, R., Portnov, M.E.,
Praasad, A., Reddix-Dugue, N., Schandler, K., Schueler, M.G., Shah, K.,

Sison, C., Stantirrop, S., Thomas, J.W., Thomas, P.J., Tsipouri, V.,
Vogl, J.U., Wetherby, K.D., Young, A. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 207486)
Green, E.D.
Direct Submission
Submitted (12-JUN-2003) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Galtersburg, MD 20877, USA
3 (bases 1 to 207486)
Green, E.D.
Direct Submission
Submitted (15-AUG-2003) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Galtersburg, MD 20877, USA
On Aug 15, 2003 this sequence version replaced gi:31621317.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@ngi.nih.gov
----- Project Information
Center project name: es1
Center clone name: 032801

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 20481 bases at least Q40
Consensus quality: 20576 bases at least Q40
Consensus quality: 206263 bases at least Q20
Insert size: 192000; agarose-fp
Insert size: 206486; sum-of-contigs
Quality coverage: 12.59x in Q20 bases; agarose-fp
Quality coverage: 11.71x in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1
430: contig of 430 bp in length
431 530: gap of unknown length
20322 20321: contig of 19791 bp in length
20322 20421: gap of unknown length
20422 99662: contig of 79241 bp in length
99663 99762: gap of unknown length
99763 106538: contig of 6776 bp in length
106539 106638: gap of unknown length
106639 113489: contig of 6851 bp in length
113490 113589: gap of unknown length
113590 143460: contig of 29871 bp in length
143461 143560: gap of unknown length
143561 149849: contig of 6289 bp in length
149850 149949: gap of unknown length
149950 151973: contig of 2024 bp in length
151974 152073: gap of unknown length

*	152074	160497: contig of 8424 bp in length
*	160498	160597: gap of unknown length
*	160598	170482: contig of 9885 bp in length
*	170483	170583: gap of unknown length
*	170583	207486: contig of 36904 bp in length

ORIGIN

Alignment Scores:

Pred. No.:	1 69e-47	Length:	207486
Score:	1210.50	Matches:	314
Percent Similarity:	26.33%	Conservative:	2
Best Local Similarity:	26.17%	Mismatches:	1
Query Match:	48.44%	Indels:	803
DB:	2	Gaps:	3

US-09-993-966-7 (1-470) X AC145175 (1-207486)

QY	154	SLuApplleThSerLeuLeuHsthrilleYrGluValValAAsSerSerValAAsHis	173
QY	177455	CAGGCAATCACCAGGCTGTGTGACACCACTCTAGAGGTGTGACTCTCTGTCAACAC	177514
QY	174	SeProHsrSerSerLyseMetLeuArgValLySleuThrValAlaProAspGlySerGln	193
Db	177515	TCCCAACATCTCAGCAAGATGTGTGGGGGTAAAGCTCACCGTGGCCCCCGATGGCAGCCAG	177574
QY	194	SeLyAsArgSerValLeuValAsnGlnAla-----	203
Db	177575	AGCAAGAGGAGCGCTCTGTTCATCAAGGCTGTGTGAGGGCTGCAGGGCTGACCTGGGAG	177634
QY	203	-----	203
Db	177635	CAATGTCTCCCATCTCAGAGGAAGAAACAGAGCCTTCCTGGAGGGGCCGGAAAGGAACTTG	177694
QY	203	-----	203
Db	177695	TGGGCTTCTGGGGCAGCTGCTTTGTGACAGAGTGGGGTTCAATGAGTTCCTTGAAGGG	177754
QY	203	-----	203

Db	17775	TAGAGGAAGGCTGGAAAGGGAGGAAAGGATTCTGGGTGAGAGGCTCATATATAGCAG	177814
QY	203	-----	203
Db	177815	AGCCATGACGCTGGATGTAAGGCTGTGTATTATAGCAGACGCGCTTCAAGCTTGAG	177874
QY	203	-----	203
Db	177875	GCCAGCTGAAAGAGGTAGAGCTGAGAGGGGAGACTGGGGAGAGGTAGAGCTAGATGG	177934
QY	203	-----	203
Db	177935	GGATGTGTCCAAGGAGGTAGCCGTGAGACCCCTTTATTTAGCCCACTCAGGAAACC	177994
QY	203	-----	203
Db	177995	TCCCAAGTGGCTTTGCAAAAGCACTGCACGCCAGCAGCCCTGGGAGGTTGGAGCAG	178054
QY	204	-----	209
Db	178055	GCGGCTCACCTGGTGTCTCTGTGCTTCTCGGGCCGAGACTCAGACCTTGCAAGGCGCAAG	178114
QY	210	ProAlaIaGluThrLysProThrGluAspLeuArgSerTyrGluLysIysGlnAlaGla	229
Db	178115	CCCCAGACAGAGACCAAGCCCACTAGAGACCTGCGAGCTGGAGAGAAAGCAGCAGCC	178174
QY	230	ProLeu-----	231
Db	178175	CCGCTCAGGTATGTGGGCAATGGTCACATAGCATATGTGAGCACCACTGTGTACTTG	178234
QY	231	-----	231
Db	178235	CTTCTGAAGATTGAGAGCAGAGATTGTTGCTCTGTGTCTGTGTTCTTCAAGGAGCCAA	178294
QY	231	-----	231
Db	178295	CACTTTGTATGCACAGACTCATTAATCAATGAAAACTTTAACAAATTAATACTAGTATGT	178354
QY	231	-----	231
Db	178355	CCACATCCTTTTATGGAACACATCTTGGGCACACAGAGACTGAGAGCCAGATTTTTGGG	178414
QY	231	-----	231
Db	178415	CTTTGAAAGGTCTCTCAGCAGTGCATCAGGCTCTGCAGCGTTGTTCTGTGCTGGA	178474
QY	231	-----	231
Db	178475	GATGACTGAATATTTCACACTAAGTAGATAAAGACCGTAAAGCCTTCCATCCCTTTAG	178534
QY	231	-----	231
Db	178535	GGTGGCGCTGCTGCATGTGGGTTATGAAAACCTTAGTTTTCAGGCTTTTGGACTTC	178594
QY	231	-----	231
Db	178595	AGAAATGTGATATGAAGAAATTATGAGCGTGTAGTACTTTCAACTGTATCTTCTAGCT	178654
QY	231	-----	231
Db	178655	GCAAGGAGAGAGGAGGAAAGAAAGAGAGATTGTTTAAAGAAAGGATTGAGTAACA	178714
QY	231	-----	231
Db	178715	ACAGTACACGCGACGACATTAAGCTGTGTATACTTACTAGTGTGCAATAGTAAGGAATG	178774
QY	231	-----	231
Db	178775	ATCGAGCTGGAATTGTACTCGCTTGGGGAAAGAAAAACAGTTCAAGCAGGCTCACAGC	178834
QY	231	-----	231

Dh 178835 CAGGTGTGGGGACAGGGCTGAGATCATAGAGGACCAGGCTGGAGATCTGGGGGA 178894
Qy 231 ----- 231
Db 178895 AGGTGTACTGTGTGTGTGGGAAGAACGAGGAGGAGCATGAGAGGTCTTGG 178954
Qy 231 ----- 231
Db 178955 AGGAGAGGGTAGGAGCCCCGAGCCCTGCAGGGGAAAAAGAAAGCCTTGGCAGCATCT 179014
Qy 231 ----- 231
Db 179015 CTCCCAACCCCTCCATACACACTGTGTAAGTACACATATGTGCACACACACTCAC 179074
Qy 231 ----- 231
Db 179075 ACACGTGTAGAGCCCATCTGGGCGTAGGGCAGGTGATGGGGTCTGCTGGCCAAACC 179134
Qy 231 ----- 231
Db 179135 CAACAGTTTAAAGGGATTCTGTGCCACACCCCTGAGAAACAGGCCAGGGGTTAGGC 179194
Qy 231 ----- 231
Db 179195 CGCTTACCTGCAGTGGGCAATATGCTTGAAAGCTTGCACTGGGCACTTTAGAAAGTT 179254
Qy 231 ----- 231
Db 179255 GACCAATTAACCTCTCAGTTTCTTTGAACCTGTTGAGCAAAACCCGTCACATCTC 179314
Qy 231 ----- 231
Db 179315 CCAGAACTGCGGGAGGTGCGGCTGTGGCTGTGTATGATCTGTCTTCATCAGGC 179374
Qy 231 ----- 231
Db 179375 AAGGAGGAATGAGGACACAGTTTGTATAGAGACTGTTCTCCACCACCCCTCTCC 179434
Qy 231 ----- 231
Db 179435 GAAGGCCAAGAGAAATGATGAGAGGATCCAGAGTTCAATTCTGGGGGCTTCTAGTTGC 179494
Qy 231 ----- 231
Db 179495 CTGTGGCTTGCCCCCACTGTGTGTGTATCCCACTCACTTGCCTCCCTGCGT 179554
Qy 232 AArgPheGlnIYAAsPseArgLeuGlnSerGlyCySerThrIshIscYValAspGlu 251
Db 179555 AGGTTCCAGGGGTGACAGCCCGCTGGAGCAGTCTGGCTGCTACCACTGCGTAGATGAG 179614
Qy 252 AasnIleGluArgArgAsnIleTYrLeuAspLeuAlaGlyIleGluAsnTYrThrSerGln 271
Db 179615 AACATGAGAGGAGAAACCACTTATATCTGCGCGGATAGAAAACCTACACGTCCAG 179674
Qy 272 PheGlyPro----- 274
Db 179675 TTTGGGCTGTGTAAAGGACTCAAGCACCTGCAATGGGCGATGAGGCGAGGCTGCTG 179734
Qy 274 ----- 274
Db 179735 GACGGGCGAGGGCGGCGCATGCAAGTGTGTCTACTGCTACCCAGGCTTGGTAAAGCAG 179794
Qy 274 ----- 274
Db 179795 CTATTATGGACAGTCAAGAACTTTTGGAGAAAGTACGTTAAAAATGGTTTCAAAAT 179854
Qy 274 ----- 274
Db 179855 TAAAGTATACATGACGAAGATCGAACAAATCCAAATATGATCTAAAGGCTTAGCACCCA 179914
Qy 274 ----- 274
Db 179915 ATGACAAACAAAGTTCTGTGCCCATCCAGGCCCAAGCCCTGTTCTTGGGGGCAAC 179974

Qy 274 ----- 274
Db 179975 CACTTGAACCTTCTCGGCTGTTTCTTCAGCAAGAACTCGGTTTCTTAAGTACAGCA 180034
Qy 274 ----- 274
Db 180035 CTGCTGTCTTATACAGCTTGACATATGTGACCTCTTGTCATTTCACTCCACAA 180094
Qy 274 ----- 274
Db 180095 CTCCCTTCTCCCTTCTTCATATGTAAATTCTCTATGAGTTACTTGTGATGTTT 180154
Qy 274 ----- 274
Db 180155 TTAGCAATATACTAGAGTTCTTTCTTGCCCAAGTACCTTAGATAGTTTTCGGGATCTC 180214
Qy 274 ----- 274
Db 180215 TGCTGCCATGTCTTCTTCTCGGCTCCCTCTCGAGAGATTTTGAAGACCTACCAAGA 180274
Qy 274 ----- 274
Db 180275 ACCAGTTCTTCCACTTAGCAGATGGGAAAGTTCCATTTGAGAGAGTTTCTGCAAGT 180334
Qy 274 ----- 274
Db 180335 AGTAGAGGAGATCAGGGCGTTGGGGGTAGCTCTGTGTTTGGAGAACTGGGGCGGGGG 180394
Qy 274 ----- 274
Db 180395 TGATGTCTGGGATATAGCGCAAGCCCAAGACACAGTAGGTGCTCAATTAATGTCTGTTG 180454
Qy 275 ----- 275
Db 180455 AATTGTTCTTAGAGCTCCCTTCGTGCGCCAGAAAGTCAAACTGCCCGGACCTCC 180514
Qy 291 AasnProThrArgSerArgSerIleGluProGluAlaIleIshIleProIshArgIlePro 310
Db 180515 AATCCCACTGATCTGCTCCCTCCATGAGCCGGAAGCCATCCACGTCACACCGAAGCCC 180574
Qy 311 GlnGlyValAspProIshSerPheIshPheLeuAspThrProIshIshValSerGlu 330
Db 180575 CAAGGCGTGAACCGGCTCTCTTCACTTCTTGACACCCCAATCGCAAGTCTCAGAG 180634
Qy 331 LeuGlnIshArgLeuArgGlyThrGlnAspGlySerIshIshPheValArgSerProIsh 350
Db 180635 CTCGAGCAAGGCTCCGGGCAACCAAGAGGAGCAAGCACTTTGAGAGTCCCCCAAG 180694
Qy 351 AlaGlnIshIshSerValGlyValGlyIshIshValIshArgGlyValIshArgAsnIshProPro 370
Db 180695 GCCCAGGGCAAGAGTGTGGGTGTGGGCACTGGGCGGAGGAGGCAAGAAACAAGCCCTC 180754
Qy 371 LeuGlyProAlaIshProAlaValSerProSerAlaIshIshIshIshIshIshIshIsh 390
Db 180755 CTGGAGCGGCAATCTCTGTGGGTGTCCCTTCAACCACTGTGCTGACCGGCTTC 180814
Qy 391 LeuProSerLeuAlaProLeuGlyIshIshIshIshIshIshIshIshIshIshIshIsh 410
Db 180815 CTCCCTCTCTTACCCCTCCGCGGCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 180874
Qy 411 GlnGlyCyshArgGlyLeuGlnAlaProLeuAlaSerGlyIshIshIshIshIshIshIsh 430
Db 180875 CAGGGCTGCGGGGCGGCGGCGAGCAACCTGCGTGTGGGCTGCGGCGGCGGAG 180934
Qy 431 HisLeuArgGluLeuProAlaLeuValIshIshIshIshIshIshIshIshIshIshIsh 450
Db 180935 CACCTGGGAGGCTGCCGCTGTGTGTGTATGAGAGCCAGGCGGCGGCTCAG 180994
Qy 451 ArgHisGluHisIshIshIshIshIshIshIshIshIshIshIshIshIshIshIshIsh 470
Db 180995 AGACATGAGCAACACACACCACTGAACATCAACACCATTAACACACACTTCTACAGACA 181054

RESULT	13				
LOCUS	HSA303140				
DEFINITION	HSA303140	16319 bp	DNA	linear	PRI 03-APR-2004
ACCESSION	AJ303140				
VERSION	AJ303140.1	GI:14488148			
KEYWORDS	LRR-containing protein, NOD2 gene.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
REFERENCE	1				
AUTHORS	Hugot, J.P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J.P., Belaiche, J., Almer, S., Tyek, C., O'Morain, C.A., Gasull, M., Binder, V., Finkel, Y., Corbell, A., Modigliani, R., Laurent-Puig, P., Gower-Rousseau, C., Macry, J., Colombel, J.F., Sabatou, M. and Thomas, G.				
TITLE	Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease				
JOURNAL	Nature 411 (6837), 599-603 (2001)				
MEDLINE	21279172				
PUBMED	11385576				
REFERENCE	2 (bases 1 to 163319)				
TITLE	Zouali, H.				
JOURNAL	Direct Submission				
FEATURES	Submitted (02-JAN-2001) Zouali H., Genetics, Fondation Jean Dausset-CEPH, 27 RUE Juliette Dodu, 75010 Paris France, FRANCE				
source	1..163319				
	Location/Qualifiers				

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q12-13"
/clone="hb87b10"
/clone_lib="CEPH Human BAC library"
98302..131994
/gene="NOD2"

join(98302..98906,106807..106912,109510..111325,
115517..115600,115821..115904,116855..118938,
121552..121635,122330..122313,124418..124501,
128745..128828,130674..131994)
/gene="NOD2"
98302..98906
/gene="NOD2"

/number=1
join(98448..98906,106807..106912,109510..111325,
115517..115600,115821..115904,116855..118938,
121552..121635,122330..122313,124418..124501,
128745..128828,130674..130746)
/gene="NOD2"
/function="putative NF-kB activator"
/note="protein of the CARD family"
/codon_start=1
/product="LIR-containing protein"
/protein_id="CAC42117.1"
/db_xref="GI:14488149"
/db_xref="GOA:O9HC29"
/db_xref="Swiss-Prot:O9HC29"
/translational=MCSQEAFOKORSOLTELLIYSGSLGSESVLDMLIMSEVLSMEHY
EGPFLILQPLSHLARRLDLPTVMKWTGWCAQCIIAAQOEADSPSFKLGCHWDPSHSI
PARDQSREPIYRIVLRSHVENKIDLAEMERGVSQTDECRILPFTFSQARLYSL
ATVAIVTENVALLDHQWAGSPOKSPATIGLEIFSTPHGLNDADQTFSTYGAETL
LCLEDITANGVALHWADVAGASPOKSPATIGLEIFSTPHGLNDADQTFSTYGAETL
GSKTLDRHLHLMAGDDFQEFELFPFCSCROLOCMALSVRTLFEHCGMPVOGEQ
DI FOLDHDPRIYVLFDGEDERKFPTFRBRHGCPATPTSQTLFFHLCMGLIKNAVA
RKVTYSRAVSNVAFIKRYITENKLGSSEOGIYLIRKHNEPVADRILIOLETS
AHDCGRAPVSMVAKCHEBLDLSGGSPKTTMTMYLIIOHTLIA TPPDSAOCL
GPSLGRRLPTLHLGKLVLMGAECCCTFYSQILOAPRPEDSDISDFIVRAGVVP
GSTVPLEHRTLTQCFPAAYIALSLADVEPALLRHLPCNGRGNSPMARLPWCIOA
SEGDSVAALLQGVFAAVNQITPAFLDGVRHMGVLAEQTESKALLRRQACAM
CLASSIKRPFESLTPAPAGSAKVHAMPGFIWLIRSLYMEGERLARAAQGANVHLHI
DTLCSVGPEFHCAALAVLOHLRPPVALQDNVSDIGVEQLCEGVCAALVDNRN

```

intron	106807.106806 /gene="NOD2" /number=1	163319
exon	106807.106912 /gene="NOD2" /number=2	315
intron	106913.109509 /gene="NOD2" /number=2	1
exon	109510.111325 /gene="NOD2" /number=3	1
intron	111326.11516 /gene="NOD2" /number=3	883
exon	115517.115600 /gene="NOD2" /number=4	1
intron	115601.115820 /gene="NOD2" /number=4	1
exon	115821.115904 /gene="NOD2" /number=5	1
intron	115905.118854 /gene="NOD2" /number=5	1
exon	118855.118938 /gene="NOD2" /number=6	1
intron	118939.121551 /gene="NOD2" /number=6	1
exon	121552.121635 /gene="NOD2" /number=7	1
intron	121636.122229 /gene="NOD2" /number=7	1
exon	122230.122313 /gene="NOD2" /number=8	1
intron	122314.124417 /gene="NOD2" /number=8	1
exon	124418.124501 /gene="NOD2" /number=9	1
intron	124502.128744 /gene="NOD2" /number=9	1
exon	128745.128828 /gene="NOD2" /number=10	1
intron	128829.130673 /gene="NOD2" /number=10	1
exon	130674.131994 /gene="NOD2" /number=11	1
ORIGIN		
Alignment Scores:		
Pred. No.:	1.06e-46	163319
Score:	1191.50	315
Percent Similarity:	26.36%	1
Best Local Similarity:	26.27%	1
Query Match:	47.68%	883
DB:	9	3

US-09-993-966-7 (1-470) x HSA303140 (1-163319)

QY	15	Glubep11ePhrSerLeuLeuH1eTh11eTyGluVal1aPseSerVal1aHn1s	173
Db	28958	CAGGACATCACAGCTTGCTGCACACCATATATAGGTGGTGAATCTCTGTCAACAC	29012.7
QY	174	SePProThiSeSer1ySmelLeuAr1yVal1yLeuThyVal1aProApG1ySerGin	193
Db	29018	TCGCCAATCTCACACAGATGCTCGGGTAAAGCTACCGTGGCCCCGATGGCACCG	29077.7
QY	194	Se1y1aAr1ySerVal1eVal1aArg1a1a-----	203
Db	29078	AGCAAGAGAGAGCGCTTGTTCATACAGCTGTGTAGGGCTGCAGGGCTGAGCTGGGAAA	29137.7
QY	203	-----	203
Db	29138	CAATGTCTCCCATCTCAGAGAAAGAACAGAGCTTCTGTGGAGGCCGGAAAGGAACCTG	29197.7
QY	203	-----	203
Db	29198	TGGGCTTTCTGGGGCAGCTGCTTTGTGACAGGTTGGGGTTCAAAATGGTCTTGAAAGG	29255.7
QY	203	-----	203
Db	29258	TAGAGGAAGGGTGGGAAGGGAGGGAAGGAATCTGGGTGAGAGGCTCATATAGCAG	29312.7
QY	203	-----	203
Db	29318	AGCCATGACAGCTGATGTAAGCTGTGTTACGGGCAGCATGGCGCTCTCAAGCTTGAG	29377.7
QY	203	-----	203
Db	29378	GCCAGCTGAAGAGAGTGAAGGTCTGAAGGGAGTCTGGGGAGAGAGTGAAGCTAGATG	29437.7
QY	203	-----	203
Db	29438	GGATGTCTCAGGGCAGATGAGCTTGAGACCCCTTATTTAGCCAGCTCAGGAAACC	29497.7
QY	203	-----	203
Db	29498	TCCCAAGTGGCCTTGCAAAAGCACTGACAGCCAGCCAGCCCTGGGGAGGGTGGAGCAG	29557.7
QY	204	-----AspLeuGlnSer1aArg	209
Db	29558	GCAAGCTCAGCTGTGTCTCTGTGCTTCTCGGGCCGGACTCAGAGCTCGACAGCCGAAG	29617.7
QY	210	ProArg1aG1uPhr1yPProThnG1uAspLeuAr1ySer1ProG1u1yLe1yGlnArg1a	229
Db	29618	CCCCGACGAGACCAAGCCCACTGAGGAACCTCGGAGCTGGGAGAAAGCAGCGAGCC	29677.7
QY	230	ProLeu-----	231
Db	29678	CCGCT-CAGGATGTGGGATGATGATGATGAGCATATGTTGAGCACAGCTGTGACT	29736.7
QY	231	-----	231
Db	29737	GCTTCTGAAGATTCAAGAGAGAAATTGTTCTGTGTGTCTCTTCAAGAGCCA	29796.7
QY	231	-----	231
Db	29797	ACACTTTGATGACAGACTCATTAATCAAAATGAAATCTTAACAATAACTACATGTAG	29856.7
QY	231	-----	231
Db	29857	TCCACATCTTTATATGAMACACTTTGGCAGACAGACTGGAGGTGACATTTTGG	29916.7
QY	231	-----	231
Db	29917	GCTTTGAAAAGTGCTCTCAGCAGTGCATCAGGCTCTGACGTGTGCTGTTCTGCTGG	29976.7
QY	231	-----	231

Db	29977	AGATGACTGAATATTTCACTAAGTAGATGAATAAAGCCGTAACACCTTCCATCCCTTTA	30036
QY	231	-----	231
Db	30037	GGGTGGGCGTGGCTGCCACGTGGGTTATGAAAACCTTAGTTTTCAGGCTTTTGGACTT	30096
QY	231	-----	231
Db	30097	CAGATGTGGATTAAGGAATTATGAGGCTGTAGTTACTTTGACCTGTATCTTTCAGC	30156
QY	231	-----	231
Db	30157	TGCAAGGAGGAAGGAGGAGAAAGAAAGAGATTGTTTAAAGAAAGGATTGAGTAAC	30216
QY	231	-----	231
Db	30217	AACAGTAACACGGCAGGCATTAAGGCTGTGTTAACCTTAGTGTGAAAATGTTAAGGAAT	30276
QY	231	-----	231
Db	30277	GATCGAGCTGATTAGTACTCGCTAGGGGAAAAAGAACAGTTCAAGCAGGCTCACAG	30336
QY	231	-----	231
Db	30337	CCAGTGTGGGACAAGGCTGAGAAATCAAGAGGACCCAGGCTTGATAGATCTGGGGG	30396
QY	231	-----	231
Db	30397	AAGTGTACTGTGTGTGTGTGGAAAGAAAGCAGAGGAGCAGATCGAAGAGTCTTGG	30456
QY	231	-----	231
Db	30457	GAGGAGAGGCTAGGAGACCCGAGCCCTGAGAGGAAAAAGAAAGCCTTGGACAGATC	30516
QY	231	-----	231
Db	30517	TCTCCCAACCCCTCATGCACTGTGTAAGTACACATGATGTGCACACACAGACTCA	30576
QY	231	-----	231
Db	30577	CACACTGTAGAGCCATCTGGGCGCTGAGGCGAGTCAATGGGGTCTGCTGGCCCAAC	30636
QY	231	-----	231
Db	30637	CCAACAGTTTAAAGGGATTTCTGTGCGCCACACACCCTGAGAAACAGGCAAGGGTTAG	30696
QY	231	-----	231
Db	30697	CCGCTTACTTCAGTGGGACATATGCTTTGAAGCCTGGCACTGGGGCATTCCTAGAAAT	30756
QY	231	-----	231
Db	30757	TGACCAATTAACCTCTACGTTGCTTTGAACCTGTTCAGAGCAAAACCCCGTCACTCT	30816
QY	231	-----	231
Db	30817	CCCAAGAGCGGGGGAGAGGTCCGGGCTGTGGGCTGTGTATGTGTGTCTTCCATGG	30876
QY	231	-----	231
Db	30877	CAAGAGGAGAAATGAGAGCACAGTTCGTATAGAGACTGTCTCCCAACCTCTCCAAA	30936
QY	231	-----	231
Db	30937	GGCCAGAGAGAAATGATGAAGGCTCAAGAGTCAATTCTGGGGGCTTCTTAGTAGCTTA	30996
QY	232	-----ATG	232
Db	30997	TGCGCTTGGCCCAACCTGGGTGTGTGTATCCCACTCACTTGGCTCCCGCCGCTAAG	31056
QY	233	PheGlnGlyIshSerArgLeuGlnGlnInsArgIYcstyYrHisGcysValIshpGluAsn	252
Db	31057	TTTCCAGGATGACAGCCGCTGGAGCAGTCTGGCTGTACACACATTTGGTATGATGAAC	31116

QY 253 ILeGIuArGrAaRnHsTYrLeuAsPLeuAlaGlyIleGIuAsnTYrTrsSerGlnPhe 272
 Db 31117 ATGAGAGGAGAAACCACTACTTATGATCTGCGGAGATAGAAACATACAGTCCCAATTT 31176
 QY 273 GlyPro----- 274
 Db 31177 GGGGCTGTAAAGGACCTCAAGCACCTTCAATGGCGATGAGGCGAGGCGTGGCTGAC 31236
 QY 274 ----- 274
 Db 31237 GGGCGAGGCGGCGCTGCGGGGTGTCTACTGCTACCCAGGCTTGGTAAGCAACTA 31296
 QY 274 ----- 274
 Db 31297 TTATGGACAGGTCAAGAAGCTTTGGAGAAAGTACGTTAAATGGTTTCAAAATTAA 31356
 QY 274 ----- 274
 Db 31357 AGTAATCATGACGAATATGAAATCAATATGATCTAAAGCTTAGACCCAAATG 31416
 QY 274 ----- 274
 Db 31417 ACAAAACAAGTTTGTGCCCCATCCAGCCCAAGCCCTGTTCCTGGGAGCAACAC 31476
 QY 274 ----- 274
 Db 31477 TTGACCTCTCTGCGCTGTTCTTCCAGCAAGAACTTCGGTTTCCCTAAGTACGACTA 31536
 QY 274 ----- 274
 Db 31537 CTGTGTCTTATACAGCTTGACATATGTGACCTTGTCAATTATTCACCCCAACTC 31596
 QY 274 ----- 274
 Db 31597 CTTTCTCCCTTCTCTTCAATGTAAATTCCTATGAGTACTTCTGTATGTTTTTA 31656
 QY 274 ----- 274
 Db 31657 GCATATACTTAAAGTTCTTCTTCCGCCAGTGCCTTAGATAGTTTTCGGGCTCTGC 31716
 QY 274 ----- 274
 Db 31717 TGCATCTGTCTTCTGCGGCTCCCTCGAGAGTATTTTGAAGACTACCAAGAAC 31776
 QY 274 ----- 274
 Db 31777 AGTTCTTCCACTTAGACAGATGGAAAGTTTCATTTGACAGAGTTTCTGCAAGCAGT 31836
 QY 274 ----- 274
 Db 31837 GAGGGGACGTCAAGGCAATTGGGGGTAGCTGTGTTTGGAGAACTGGGGCGGGGTGA 31896
 QY 274 ----- 274
 Db 31897 TGTCTGGGATATAGCGCAAGCCGACGACACAGTAGGTGCTCATTAATGTCTTGAAT 31956
 QY 275 -----GlySerProSerValAlaGlnlySerGluLeuProProAlaGlnSerAsn 291
 Db 31957 TGGTTCTTAGGCTCCCTTCGCTGCGCCAGAAAGTCAAAAGTGGCCCCCGGCACTTCAAT 32016
 QY 292 ProThrArgSerArgSerHisGluProGlnAlaAlaHisIleProHisArgIleValSerGln 311
 Db 32017 CCCACTGATCTCCCTCCCATGACCGGAAAGCCATCCACATCCACACCGAAAGCCCCAA 32076
 QY 312 GlyValAspProAlaSerPheHisPheLeuAspThrProIleAlaIleValSerGluLeu 331
 Db 32077 GGGGTGAGCCCGGCTCTTCACTTCTTACACCCCAATCGCCAAAGTCTCAGACTC 32136
 QY 332 GlnGlnArgLeuAlaGlyThrGlnAspGlySerIleHisPheValAspSerProIleAla 351
 Db 32137 CAGCAAGGCTCCGCGGAGCAACCAAGACGGGAGCAAGCACTTGTGAGGTCCCGCAAGGCC 32196

QY 352 GlnGlyLeuSerValGlyValGlyHisValAlaArgGlyAlaArgAsnlyProProLeu 371
 Db 32197 CAGGCCAAGAGTGTGGGTGTGGGCCACGTGGCCAGAGGGGCAAGAAACAGCCCTCTG 32256
 QY 372 GlyProAlaIleProAlaValSerProSerAlaHisLeuAlaAlaSerProAlaLeu 391
 Db 32257 GGAACCGGCATCTCCCTCGGTGTCCCTCCGCCACCTGCTTCAGAGCCGCGCTCTC 32316
 QY 392 ProSerLeuAlaProLeuGlyHisIleValSerHisIleValArgAlaIleValSerGln 411
 Db 32317 CCTCTCTAGCCCCCTTGGGACAAAGACCAAGACCCGACCAAGAGACCCAGCAG 32376
 QY 412 GlyCysArgGlyLeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlyArgIleHis 431
 Db 32377 GGTGCGGGGCTGTGAGGACCACTGGCTTCAAGTGGCTCTCTGGGGCGGAGCAC 32436
 QY 432 LeuArgGluLeuProAlaLeuValIleValTYrGluSerGlnAlaGlyGlnProValGlnArg 451
 Db 32437 CTGCGGAGCTGCGCCCTTGGTGTGTAGAGAGCCAGGCCGCGGACGCGTCCAGAGA 32496
 QY 452 HisGluHisHisHisHisGluHisHisHisHisHisHisHisHisHisHisHisHis 470
 Db 32497 CATGAGCACCAACCAACATGAAATCACACCATTTACCACTTACCAAGCA 32553
 RESULT 14
 AC007608 169773 bp DNA linear PRI 19-MAR-2003
 LOCUS
 DEFINITION Homo sapiens chromosome 16 clone RP11-401P9, complete sequence.
 AC007608
 VERSION AC007608.7 GI:29124038
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 169773)
 REFERENCE
 AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los
 Alamos National Laboratory.
 TITLE Direct Submission
 JOURNAL Unpublished
 AUTHORS 2 (bases 1 to 169773)
 Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
 Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
 Bryant,J., Tesmer,J., Meincke,L., Lonmize,J., White,S., Tatum,O.,
 Campbell,C., Rawcett,J., Maltbie,M., Bussed,M., Sutherland,R.,
 McMurtry,K., Han,C. and Deaven,L.
 DIRECT SUBMISSION
 Submitted (20-MAY-1999) Center for Human Genome Studies, DOE Joint
 Genome Institute, Los Alamos National Laboratory, MS M888, Los
 Alamos, NM 87545, USA
 3 (bases 1 to 169773)
 REFERENCE
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (04-JAN-2002) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 4 (bases 1 to 169773)
 REFERENCE
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (11-JAN-2002) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 5 (bases 1 to 169773)
 REFERENCE
 AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los
 Alamos National Laboratory.
 TITLE Direct Submission
 JOURNAL Submitted (19-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 On Mar 19, 2003 this sequence version replaced gi:18129388.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.igi.doe.gov
 Finishing Completed at Stanford Human Genome Center and Los Alamos
 National Laboratory
 www.sngc.stanford.edu
 Quality: Phrap Quality >=40 100% of Sequence;

Db 79673 CCCAAGAGCTGCGGGAGAGTCGGGCTGTGGGCTGTGGTCTATGATGTCTGTCTTCATCGG 79614
 QY 231 ----- 231
 Db 79613 CAAGAGGAAATGAGAGCA GATTTCGTATAGAGACTGTTCTCC CACCCTCTCCAA 79554
 QY 231 ----- 231
 Db 79553 GGGCAAGAGAAATGATAGAGGCTCAGAGTTCA TTTGCGGGGCTTCTTACAGCTTA 79494
 QY 232 ----- Arg 232
 Db 79493 TGGCCTTGCCCCCTGATGATGATGATTTATCCCACTTCTCCCTGCGGTAGG 79434
 QY 233 PheGlnGlyAspSerAlaGluGlnSerGlyCysValAspGln 252
 Db 79433 TTCGAGGATACAGCCCTGAGAGCTGGCTGTACCA CATTGCGATGATGAAAC 79374
 QY 253 IleGluArgArgAsnHisTyrLeuAspLeuAlaGlyIleGluAsnTyrThrSerGlnPhe 272
 Db 79373 ATCGAGAGGAAACCACTTATGATCTCGCGGGATAGAAACTACGTCCTCAATTT 79314
 QY 273 GlyPro ----- 274
 Db 79313 GGGCTGTGAAGGACTCAAGCACTTGAATGGCGATGAGGCA GGGCGTGGCTGAC 79254
 QY 274 ----- 274
 Db 79253 GGGCGAGCGGGCCGTGGGGGTGTTCAC TGTCTACCCAGGCTTGGTAAGCAACTA 79194
 QY 274 ----- 274
 Db 79193 TTATGGGACAGGTCAAAAGACTTCTTGAGAAAGTACGTTAAAATGTTTCAAAATTA 79134
 QY 274 ----- 274
 Db 79133 AGTATATCATGACGAATATGCAAAATCCATATGATCTAAAGGCTTAGACCCCAATG 79074
 QY 274 ----- 274
 Db 79073 ACAAACAAGTTTGTGCCCATCCAGCC CACCAGCCGTGTCCGGGGCAACAC 79014
 QY 274 ----- 274
 Db 79013 TTTGACCTCTCCGGCTGTTTCTTCCAGCAAGAACTCGGTTCCCTAAGTACAGCACTA 78954
 QY 274 ----- 274
 Db 78953 CTGTGCTTTATCAGCCTTGACATATGTGACTCTTGTCAATTTCAC TCCACAATC 78894
 QY 274 ----- 274
 Db 78893 CCTTCTCCCTTCTCTTCATATGTTAA TTTCTCTATGAGTTACTTCTGTATGTTTTTA 78834
 QY 274 ----- 274
 Db 78833 GCATATACTTAGAGTTCTTCTTGCCCCCACTGA CTTAAGTAGTTTTCGGGCTCTGTC 78774
 QY 274 ----- 274
 Db 78773 TGCCATCTGTCTTCTGCGCTCCTCTCGAGAGTATTTTGAAGACTTACCAAGAAC 78714
 QY 274 ----- 274
 Db 78713 AGTTCCTCCACTTAGAGATGGAAAGTTCATTTGAGAGATTTTCTGCAAGCAGT 78654
 QY 274 ----- 274
 Db 78653 GAGGGGACAGTCAGGCAATGGGGGTAGCTCTGTGTTTGAGAACTGGGGCGGGGGTGA 78594
 QY 274 ----- 274
 Db 78593 TGTCTGGGATATAGCGCAAGCCCAAGCA CACAGTAGGTGCTCATTAATGTCTGTGAAT 78534

QY 275 -----GlySerProSerValAlaGlnIleSerGlyLeuProProArgThrSerAsn 291
 Db 78533 TGGTCTTACAGCTTCCCTTCCGTGGCCAGAACTCAGAACTGCCCCCGCACTTCCAT 78474
 QY 292 ProThrArgSerArgSerHisGluProGluAlaIleHisIleProHisArgIleProGln 311
 Db 78473 CCCACTGATCTGCTCCCATGAGCCGGAAGCATCATCCATCCACACGAAAGCCCA 78414
 QY 312 GlyValAspProAlaSerPheHisPheLeuAspThrProIleAlaIleValSerGlyLeu 331
 Db 78413 GGGGTGACCCGGGCTCTTCCACTTCTTGACACCCCAATTCGCAAGGCTTCAAGACTC 78354
 QY 332 GlnGluArgLeuArgGlyThrGlnAspGlySerGlyHisIlePheValArgSerProIleAla 351
 Db 78353 CAGCAAGCTCCGGGGACCCAGACCGAGAGCAAGCACTTTGTGAGTCCCCCAAGGCC 78294
 QY 352 GlnGlyIleSerValGlyValGlyHisValAlaArgGlyValaArgAsnIleProIleu 371
 Db 78293 CAGGGCAAGAGTGTGGGTGTGGGCCAGTGGCCAGAGGGCAAGAAACAAGCCCTCTG 78234
 QY 372 GlyProAlaIleProAlaValSerProSerAlaHisIleLeuAlaAspProAlaLeu 391
 Db 78233 GAACCCGCAATCCCTGTGGGTGTCCTCCGACCTGCTGCTGAGCCCGGCTCTC 78174
 QY 392 ProSerLeuAlaProLeuGlyHisIleValSerGlyValProValLeuGlyArgIleHis 411
 Db 78173 CCTCTCCCTACCCCTCCCTGGGCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 78114
 QY 412 GlyCysArgGlyLeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlyArgIleHis 431
 Db 78113 GGTGCGGGGGCTGCAAGGCAACACTGGCTCTCAGTGGCTCTGCTGGGGCGGAGAC 78054
 QY 432 LeuArgGlyLeuProAlaLeuValValIleGlyIleSerGlnAlaGlyGlnProValGlnArg 451
 Db 78053 CTGCGGAGCTGCGCCGCTTGTGTGTATGAGAGCAAGCCGCGGACGCGGTCCAGAGA 77994
 QY 452 HisGluHisHisHisHisHisGluHisHisHisHisHisHisHisHisHisHisHisHis 470
 Db 77993 CATGAGCACCAACACACATGAACTACCACTTACCACTTCTTACAGACA 77937
 RESULT 15
 AC007334 170791 bp DNA linear PRI 13-MAR-2003
 LOCUS AC007334
 DEFINITION Homo sapiens chromosome 16 clone R11-147B17, complete sequence.
 AC007334
 VERSION AC007334.8 GI:28933524
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 170791)
 DOE Joint Genome Institute, Stanford Human Genome Center and Los
 Alamos National Laboratory.
 TITLE Direct Submission
 JOURNAL Unpublished
 AUTHORS Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
 Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
 Bryant,D., Tesmer,J., Meincke,L., Lonmire,J., White,S., Tatun,O.,
 Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,
 McMurry,K., Han,C. and Deaven,L.
 DIRECT SUBMISSION
 Submitted (21-APR-1999) Center for Human Genome Studies, DOE Joint
 Genome Institute, Los Alamos National Laboratory, MS M888, Los
 Alamos, NM 87545, USA
 3 (bases 1 to 170791)
 DOE Joint Genome Institute.
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Submitted (24-AUG-2001) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 4 (bases 1 to 170791)
 REFERENCE

AUTHORS

DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.

TITLE

Direct Submission
Submitted (13-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA

JOURNAL

On Mar 13, 2003 this sequence version replaced gi:15284265.
Draft Sequence Produced by DOE Joint Genome Institute

COMMENT

www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www.abgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
Location/Qualifiers

FEATURES

Source

1. 170791
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-147B17"

ORIGIN

Alignment Scores:

Prod. No.: 1.12e-46 Length: 170791
Score: 1191.50 Matches: 315
Percent Similarity: 26.36% Conservative: 1
Best Local Similarity: 26.27% Mismatch: 1
Query Match: 47.68% Indels: 883
DB: 9 Gaps: 3

US-09-993-966-7 (1-470) x AC007334 (1-170791)

QY 154 GUAAPLLEthSerLeuNH1stHr1LeTyGLuValaJapSerSerValaNH1s 173
Db 31077 CAGGACATCACAGCTGCTGCAACCATATAGAGGTGGAGCTCTCTGCAACAC 31018
QY 174 SerProthSerSerlySMelLeuArgVallyLeuThrVala1A1PtoAArglySerGln 193
Db 31017 TCCCAACATCCAGCAGATGCTGCGGTAAAGCTCAACGTCGCCGCCGAGCGCAG 30958
QY 194 SerlyArgSerVallyLeuValaNH1A1A- 203
Db 30957 AGCAAGAGAGAGCTCTTGTCAATCAGGCTGTCAGAGGCTGAGACCTGGGAAA 30898
QY 203 ----- 203
Db 30897 CAATGCTCTCCATCTCAGGAAGGAAGACCTTCTGGGAGGCGGAAAGGAACCTG 30838
QY 203 ----- 203
Db 30837 TGGGCTTTCTGGGCAAGCTCTTTGTGACAGAGTGCGGTTCAATGGGCTCTTGAAGG 30778
QY 203 ----- 203
Db 30777 TAGGAGAGAGGTGGGAAGGGAAGGAGATTCTGGGTGAGAGCCTCATATAGCAG 30718
QY 203 ----- 203
Db 30717 AGCCATGACAGTGAAGTGAAGGCTGTTCACGGGCAAGCATGGGCTTGGAAGCTTGAG 30658
QY 203 ----- 203
Db 30657 GCCAGCTTGAAGAGAGGTGAGGCTGGAAGGAGAGTCTGGGAGAGAGGTAGAGCTAGATGG 30598
QY 203 ----- 203
Db 30597 GGATGTGTCCAGGGCAGGTAGCCTTGAGACCCCTTTATTAGCCCCAGCTCAGGGAACCC 30538
QY 203 ----- 203
Db 30537 TCCCAAGTGGCTTGCAAGAGAGCTGCAAGCAGGAGCCCTGGGAGAGGTGGAGACAG 30478
QY 204 ----- 204
-----AspLeuGlnSerAlaArg 209

Db 30477 GCAGCTACCTGGTGTCTCTGCTTCTGCGGCCGAGCTCAGACCTGCAAGGCGCAAGG 30418
QY 210 ProArgAlaGluThrlyPProthGluAspLeuArgSerTyrGluArgAla 229
Db 30417 CCGGACAGAGACCAAGCCCACTGAGGACCTGGGAGCTGGGAGGAAGAGCAGAGGCC 30358
QY 230 ProLeu----- 231
Db 30357 CCGCT-CAGGTATGTGGCATGTGTGACATGAGCATATGTGAGACCAAGCTGTACT 30299
QY 231 ----- 231
Db 30298 GCTTGAAGGATTCAGAGCAGGAAGTTTGTCTGTGTCTGTGTCTTTCAGGAGCCA 30239
QY 231 ----- 231
Db 30238 ACACTTTGATGCACAGACTCATTAATCAATGAAATCTTAACAATTAATACATGTAG 30179
QY 231 ----- 231
Db 30178 TCCACATCTTTATGGAACACATCTTGGCAGACAGACTGGAGTCAACATTTTGG 30119
QY 231 ----- 231
Db 30118 GCTTTGAAAGTGTCTCTCAGCAGTGCATAGGCTGTGACGTGTGCTGTTCTGCTGG 30059
QY 231 ----- 231
Db 30058 AGATGACTGAATATTCACACTAATGATGAAGAACCGTAACACCTCCATCCCTTA 29999
QY 231 ----- 231
Db 29998 GGGTGGCTGTGCTGCCACGTGGGTATGAAACTTTAGTTTCAAGCTTTTGGACTT 29939
QY 231 ----- 231
Db 29938 CAGATGTGATGAATGAATTAATGAGCTGTAGTACCTTTACCTGTACTCTTCAAC 29879
QY 231 ----- 231
Db 29878 TGCAAGGAGGAAGGAGGAAAGGAAGAGATTGTTAAAGAAAGGATTGAGTAAAC 29819
QY 231 ----- 231
Db 29818 AACGTAACAGGCAAGCATAGGCTGTGTTAACTTCTAGTGTCAATGTAAGGAAT 29759
QY 231 ----- 231
Db 29758 GATCGAGCTGATTAGTACTCGCTAGGGGAAAGAAAAACAGTTCAAGAGGCTCACAG 29699
QY 231 ----- 231
Db 29698 CAGGTGTGGGACAGGGCTGGAATTCAGAGGACCCAGGCTGGCTAGAGATCTGGGGG 29639
QY 231 ----- 231
Db 29638 AAGGTGTAAGTGTGTGTGGGAAGAACCAAGAGAGAGAGATGGAAGAGTCTTG 29579
QY 231 ----- 231
Db 29578 GAGAGAGGCTAGGAGCCCGAGCCCTGACAGGGAAAAAGAAAGCTTGGCAGATC 29519
QY 231 ----- 231
Db 29518 TCTCCCACTTCATGACACTGTGTAGTACACATGATGATGTGACACACACTCA 29459
QY 231 ----- 231
Db 29458 CACATGTGAGGCCATCTGGGCGGTAGAGGAGGATCATGGGTGTGCTGGCCCAACC 29399
QY 231 ----- 231

D	29398	CCAACAGTTTAAAGGGAATTGTGCTGCCACACCCCTGAGAAACCAAGGCGAAGGTTAG	29339
QY	231	-----	231
D	29338	CCGCCTACCTGCAGTGGGACATATGAGCTTGAAGCTGGCACTGGGGCATTCCTGAAGT	29279
QY	231	-----	231
D	29278	TGACCAATTAACCTCTCAGCTTTCCTTTGAACCTGTTCAAGACAAAACCCGTCACCTCT	29219
QY	231	-----	231
D	29218	CCCAAGAACTGCCGGGAGGTCGGGCTGTGGGCTGTGGTCTATAGTCTGTCTTCATCGG	29159
QY	231	-----	231
D	29158	CAAGAGGAAATGAGAGCACAGTTGCTATAGAGACGTGTCCTCCCACTTCACAA	29099
QY	231	-----	231
D	29098	GGCCAAAGAAATGAAATGAAAGGTCAGAGTTCATTCGGGGGCTTCCTAGACCTA	29039
QY	232	-----Arg	232
D	29038	TGCGCTTCCCCCACTCGTGTGTGTGTTATCCCACTCACTTGCTCCCTCCCGTAGG	28979
QY	233	PhcglngIyAspSerArgLeuGlnGlnInserGlyCySerPthiShiCyValAspGluAsn	252
D	28978	TTCCAGGGTGACAGCGGCTCGAGGACGTCTGCTGTACCAACATTTGGCTAGATGAAAC	28919
QY	253	IIeGluArgArgAsnHisIleTyrLeuAspLeuAlaGlyIleGluAsnTyrThiSerGlnPhe	272
D	28918	ATCGAGGAGGAAACCACTACTAGATCTCGCCGGATGAGAAAATPACACGTCCTCAATTT	28859
QY	273	GIYPro-----	274
D	28858	GGGCTCTGTAAAGGACTCAAGCACCTCTGCAATGGCGAATGAGGGCAGGGCGTGGCTGGAC	28799
QY	274	-----	274
D	28798	GGGCGAGGGGGCGGTCGGGGTGATGTTCATGCTACCCAGGCTTCGGTAAAGACAATA	28739
QY	274	-----	274
D	28738	TTATGAGCAGAGTTCAAAGACTCTCTTGAGAAAGTGACGTTAAATGTTCAAAATTTAA	28679
QY	274	-----	274
D	28678	AGTATATCATGCACAGAAATGAAACAAATTCAAATAGTATCTAAAGGCTTAGACCCCAATG	28619
QY	274	-----	274
D	28618	ACAAACAACAGTTTGTGTGCCCATCCAGGCCACCCAAGCCGTGTTCCGGGGGCAACAC	28559
QY	274	-----	274
D	28558	TTTGAACCTCTCTGAGCTGTGTTCTTCCAGCAAGAACTCGGTTTCCTAAGTACAGACTA	28499
QY	274	-----	274
D	28498	CTGTGTCTTATACAGCCTTGACATAGTGAACCTCTTGCTCAATTTCACTCCCAACAATC	28439
QY	274	-----	274
D	28438	CCTTCTGCCCTTCTCTTCAATAGTTTAATTCCTCTATAGATTACTCTGTATGTTTTA	28379
QY	274	-----	274
D	28378	GCAATATACTTAAAGATTCTTCTTGTGCCCGAGTACCTTAGATAGTTTTCGGGGTCTCTGC	28319
QY	274	-----	274
D	28318	TGCGATCTGTCTTCTCGGCTCCCTCGAGAGTATTTTGAAGACCTACCAAGAAC	28259

QY	274	-----	274
Db	28258	AGTTCTCCACTTNGACAGATGGGAAAGTTCCATTTCAGAGAGTTTTCCTGCAGAGCACT	28199
QY	274	-----	274
Db	28198	GAGGGGCACTCAGGGCATTGGGGGGGTAGCTCTGGTTTGGAGAACTGGGGGGGGGGGTGA	28139
QY	274	-----	274
Db	28138	TGTCGGGTATAGCGCAAGCCCGACACACAGTAGTGCTCAATTAATGTCTGTGAAT	28079
QY	275	-----GlySerProSerValAlaGlnLysSerGluLeuProProArgThrSerAsn	291
Db	28078	TGGTTCCTAGAGCTCCCTCCCTCCGTGGCCGAGAGTAAAGCTGCCCCCGCACCTCCAAAT	28019
QY	292	ProThrArgSerArgSerHisGluProGluAlaIleHisIleProHisArgLysProGln	311
Db	28018	CCCACTCCATCTCGTCTCCCATAGAGCCGGAAGCCATCCATCCACACCGAAAGCCCAA	27959
QY	312	GlyValAspProAlaSerPheHisPheLeuAspThrProIleAlaLysValSerGluLeu	331
Db	27958	GAGCGTGAACCCGGCTCCTCTCCACTTCTTGACACCCCAATGCCAAGGTCTCAGAGCTC	27899
QY	332	GlnGlnArgLeuArgLysThrGlnAspGlySerLysHisPheValArgSerProLysAla	351
Db	27898	CAGCAACCGCTCCCGGGGACCCAGAGACGGAGCAAGCACTTGTGAAGTCCCCCAAGGCC	27839
QY	352	GlnGlyLysSerValGlyValAlaGlyHisValAlaArgGlyValArgAsnLysProProlen	371
Db	27838	CAGGGCAAGAGTGTGGTGTGGGCCACGTGGCCAGAGGGGCAAGAAACAAGCCCCCTCTG	27779
QY	372	GlyProAlaIleProlAlaValAspProSerAlaHisLysValAlaSerProAlaLeuLeu	391
Db	27778	GGACCCCGCATCCCTGCGGTGTCCCCCTCCGCCACCTGGCTGCCAGCCCGGCTCTCTC	27719
QY	392	ProSerLeuAlaProlLeuGlyHisIleLysLysHisLysHisArgAlaLysGluSerGlnGln	411
Db	27718	CCCTCCCTAGCCCCCTCGGGGACAAAGACCAAGCACCGAGCCAAAGAGAACCGACGCG	27659
QY	412	GlyCysArgGlyLeuGlnAlaProlLeuAlaSerGlyGlyProValLeuGlyArgGlnHis	431
Db	27658	GGCTCCCGGGGCTTCAGGACCACTGGCTCAGAGTGGCCCTGTCTCTGGGGCGGAGCAC	27599
QY	432	LeuArgGluLeuProlAlaLeuValValTyrGluSerGlnAlaGlyGlnProValGlnArg	451
Db	27598	CTGGCGGAGCTGCGCGCTTGGTGTGTATGAGACCGAGCGCGGAGCGCGGTCCAGAGA	27539
QY	452	HisGluHisHisHisHisHisGluHisHisHisHisHisGlyHisHisPheHisGlyHis	470
Db	27538	CATGAGCACACACACACATGAACATCACACCACTTACACCACTTCTTACACGAGA	27482

Search completed: December 30, 2004, 06:25:17
Job time : 8483 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 30, 2004, 09:35:26 Search time 931 Seconds

(Without alignments)
2851.429 Million cell updates/sec

Title: US-09-993-966-7

Perfect score: 470
Sequence: 1 MGLKSKPAVCKKRESPEG.....RHEHHHHHHHHHHFYQT 470

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4176236 seqs, 2824127955 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8342757

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xip
-Q/cgnt1/USPTO.spool.p/US0993966/runat.29122004.130830.9445/app.query.fasta.1.647
-DB=Published Applications NA -QFMT=fastap -SUFFIX=01fg.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.csl -LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptco -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US0993966 @CGN 1.1 912 @runat.29122004.130830.9445
-NCPU=6 -ICPU=3 -NO MAP -LARGESUBV -NEG SCORES=0 -WMT -DSBLOC=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=10 -THRAPDS=1 -XGAP=60 -XGAPEXT=60
-Fgapop=6 -Fgapext=7 -Ygapop=60 -Ygapext=60 -DELop=6 -DELext=7

Database: Published Applications NA:*

1: /cgnt2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgnt2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgnt2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgnt2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
5: /cgnt2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*
6: /cgnt2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgnt2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgnt2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgnt2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgnt2_6/prodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgnt2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgnt2_6/prodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgnt2_6/prodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgnt2_6/prodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgnt2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgnt2_6/prodata/1/pubpna/US10D_PUBCOMB.seq:*
17: /cgnt2_6/prodata/1/pubpna/US10E_PUBCOMB.seq:*
18: /cgnt2_6/prodata/1/pubpna/US10F_PUBCOMB.seq:*
19: /cgnt2_6/prodata/1/pubpna/US11_NEW_PUB.seq:*
20: /cgnt2_6/prodata/1/pubpna/US60_NEW_PUB.seq:*
21: /cgnt2_6/prodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	470	100.0	1448	13	US-10-087-192-179	Sequence 179, App
2	470	100.0	1786	10	US-09-993-966-1	Sequence 1, Appl
3	470	100.0	1859	10	US-09-993-966-5	Sequence 5, Appl
4	196	41.7	2379	10	US-09-764-891-5B02	Sequence 5902, Ap
5	196	41.7	2379	10	US-09-764-891-5304	Sequence 178, Ap
6	196	41.7	106315	13	US-10-087-192-178	Sequence 178, Ap
7	195	41.5	590	10	US-09-993-966-21	Sequence 21, Appl
8	138	29.4	427	10	US-09-918-999-8686	Sequence 8686, Ap
9	49	10.4	148	10	US-09-993-966-18	Sequence 18, Appl
10	49	10.4	1401	9	US-09-730-989-1	Sequence 1, Appl
11	49	10.4	1401	10	US-09-993-966-6	Sequence 6, Appl
12	49	10.4	1416	10	US-09-993-966-2	Sequence 2, Appl
13	49	10.4	2807	13	US-10-087-192-176	Sequence 176, App
14	46	9.8	1950	13	US-10-016-157A-126	Sequence 126, App
15	45	9.6	598	15	US-10-029-386-7990	Sequence 7990, Ap
16	44	9.4	133	15	US-10-029-386-21690	Sequence 21690, A
17	44	9.4	134	15	US-09-993-966-14	Sequence 14, Appl
18	42	8.9	128	10	US-09-993-966-20	Sequence 20, Appl
19	36	7.7	639	15	US-10-243-552-231	Sequence 231, App
20	35	7.4	107	10	US-09-993-966-16	Sequence 16, Appl
21	32	6.8	96	10	US-09-993-966-17	Sequence 17, Appl
22	32	6.8	98474	13	US-10-087-192-175	Sequence 175, App
23	29	6.2	554	15	US-10-029-386-7615	Sequence 7615, Ap
24	28	6.0	179	15	US-10-029-386-21315	Sequence 21315, A
25	27	5.7	85	10	US-09-993-966-19	Sequence 19, Appl
26	22	4.7	67	10	US-09-993-966-15	Sequence 15, Appl
27	22	4.7	566	13	US-10-027-633-266931	Sequence 266931,
28	22	4.7	566	15	US-10-027-633-266931	Sequence 266931,
29	22	4.7	2724	9	US-09-833-381-1214	Sequence 1214, Ap
30	22	4.7	2724	9	US-09-833-381-1215	Sequence 83, Appl
31	17	3.6	1817	9	US-09-822-849A-83	Sequence 8220, Ap
32	17	3.6	1959	18	US-10-723-860-5220	Sequence 152, Appl
33	17	3.6	1966	9	US-09-764-868-152	Sequence 37, Appl
34	17	3.6	73028	17	US-10-322-696-37	Sequence 12063, A
35	15	3.2	674	13	US-10-027-633-12063	Sequence 12063, A
36	15	3.2	674	15	US-10-027-633-12063	Sequence 162722,
37	14	3.0	901	13	US-10-027-633-162722	Sequence 162722,
38	14	3.0	901	15	US-10-027-633-162722	Sequence 676, App
39	13	2.8	488	9	US-09-783-590-676	Sequence 13, Appl
40	11	2.3	34	10	US-09-993-966-13	Sequence 228, App
41	10	2.1	574	9	US-09-864-761-228	Sequence 17051, A
42	10	2.1	669	9	US-09-864-761-17051	Sequence 17, Appl
43	9	1.9	27	9	US-09-730-989-17	Sequence 241, App
44	9	1.9	296	15	US-10-317-444-241	Sequence 242, App
45	9	1.9	236	15	US-10-317-444-242	

ALIGNMENTS

RESULT 1
US-10-087-192-179
Sequence 179, Application US/10087192
Publication No. US20020182586A1
GENERAL INFORMATION:
APPLICANT: Morris, David W.
APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
FILE REFERENCE: 529452000122
CURRENT APPLICATION NUMBER: US/10/087,192
CURRENT FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 2059
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 179
LENGTH: 1448
TYPE: DNA

ORGANISM: Homo sapiens
US-10-087-192-179

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	470.00	470	0	0	0	0
Percent Similarity:	100.00%					
Best Local Similarity:	100.00%					
Query Match:	100.00%					
DB:	13					

US-09-993-966-7 (1-470) x US-10-087-192-179 (1-1448)

```

QY 1 MetGlyLeuHisSerLysProAlaValCysLysArgArgGlySerProGly 20
Db 39 ATGGGGAACTTCACTCCAGCCGGCCGGTGTGCAAGCGCAAGAGCCCGAAGGT 98
QY 21 AspSerPheAlaValSerAlaAlaTrpAlaArgLysGlyIleGluGluTrpIleGlyArg 40
Db 99 GACAGCTTCCCGTGAAGCGCTGGCTGGAGGCGATCGAGAGTGGATCGGAGA 158
QY 41 GlnArgCysProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGly 60
Db 159 CAGCGCTGCCCGGGCGGTGTCTCGGAGACCCGACAGCTGGCGTTGGCGGACCATAGGC 218
QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgPheThrLeuSerGluGluGlu 80
Db 219 CGAAGCACCCGGAGGCTGTGGGCGAGCTGTGAGACACAGCTCAGCGAGAAAGAG 278
QY 81 AspAspPheArgLeuGluValAlaLeuProProGlyLysThrAspGlyLeuGlySerGly 100
Db 279 GAGGACTTTCGGTGGAGAGTGGCCCTGCTCTGAGAAAGACTAGCGGGCTGGCGAGGA 338
QY 101 AspGlyLysLysMetGluArgValSerGluProCysProGlySerLysLysGlnLeuLys 120
Db 339 GATGAGAAAGAAATGAGAGAGTGAAGAAACCTCGCCAGCTTCCAAAGACAGCTGAG 398
QY 121 PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGlnGluTrpThrPhe 140
Db 399 TTGTGAGAGCTTCCAGTCCAGCTGTCCATGAGAGAGACAGCCCGCAGAGTGAACCTTC 458
QY 141 ThrLeuTyraAspPheAspAsnAsnGlyLysValThrArgLysPheIleThrSerLeuLeu 160
Db 459 ACCGTGATACCTTTGACCAACAGGAGGTCAACCGAGAGGACATCACAGCTTTCGTG 518
QY 161 HisThrIleTyrgluValAlaAspSerSerValAsnHisSerProThrSerSerLysMet 180
Db 519 CACACCATCTATGAGTGTGTGACTCTCTGTCAACCACTCCCAACATCCAGCAGAGATG 578
QY 181 LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysAspSerValLeuVal 200
Db 579 CTGGGGTAAAGCTCACCGTGGCCCCCGATGGCAGCCAGACCAAGAGAGAGGCTTGTG 638
QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgLysGluThrLysProThrGluAspLeu 220
Db 639 AATCAGGCTACCTGCAAGGCGCAAGGCCCGGAGCAGACAGCAAGCCCATGAGAGACTG 698
QY 221 ArgSerTrpGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
Db 699 CGGAGCTGGGAGAAAGAGCAGCGAGCCCGCTCAGTTCAGAGGTGACAGCCGCTGAG 758
QY 241 GlnSerGlyCysTrpHisCysValAspGluAsnIleGluArgArgAsnHisTrpLeu 260
Db 759 CAGTCTGCTGCTACCAACATGCTGATGAGAAACATCGAGAGAAACCACTACTTA 818
QY 261 AspLeuAlaGlyIleGluAsnTrpThrSerGlnPheGlyProGlySerProSerValAla 280
Db 819 GATCTGCCCGGAGTAAAGAAATACAGCTCCCAATTGGGCTGCTCCCTTCCGTGGCC 878
QY 281 GlnLysSerGluLeuProProArgThrSerAsnProThrArgSerArgSerHisGluPro 300
Db 879 CAGAGTCAAGAACTGCCCGCCGACCTCCATCCCATCGATCTGCTCCCATGAGACCG 938

```

```

QY 301 GlnAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPhe 320
Db 939 GAGGCAATCCACATCCCAACACCCAAAGCCCAAGGGGTGAGACCCGCTCTCCACTTC 998
QY 321 LeuAspThrProIleAlaLysValSerGluLeuGlnArgLeuArgLysGlyThrGlnAsp 340
Db 999 CTTGACACCCCAATCCCAAGGTCTCAGAGCTCAGACCAAGCTCCGGGCGACACAGGAC 1058
QY 341 GlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValGlyHis 360
Db 1059 GGGAGCAAGCACTTGTGAGGTCCCAAGGCCCAAGGCAAGAGTGTGGTGGCCAC 1118
QY 361 ValAlaArgGlyValaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
Db 1119 GTGGCCAGAGGGGCAAGAAACAGCCCTCTTGAGACCCGCACTCCCTGGGTGTCCCC 1178
QY 381 SerAlaHisLeuAlaAspProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
Db 1179 TCCGCCACCTGCTCCAGCCCGGCTCTCCCTCCCTAAGCCCTCCGGGCAACAG 1238
QY 401 LysHisLysHisArgAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
Db 1239 AAGCACAAGCACCTGAGCCAAAGAGACCAAGAGGCTGCCGGGCTTGAAGCACTG 1298
QY 421 AlaSerGlyGlyProValLeuGlyArgGluHisLeuArgGluLeuProAlaLeuVal 440
Db 1299 GCTTCAGTGGCCCTGTCTCGGGCGGAGACCTCGGGAGCTGGCGCTGTGTG 1358
QY 441 TyrGluSerGlnAlaGlyLysProValGlnArgHisGluHisHisHisHisGluHis 460
Db 1359 TATGAGAGCAGGCGCGGCGAGCTGTCCAGAGACATGAGCACCACCACTGAACAT 1418
QY 461 HisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 470
Db 1419 CACCACCATTAACCACTTTCACAGACA 1448

```

RESULT 2

```

US-09-993-966-1
; Sequence 1, Application US/09993966
; Publication No. US20030186232A1
; GENERAL INFORMATION:
; APPLICANT: ROMAN, MICHAEL
; TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF MKD PROTEIN,
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USBS THEREOF
; FILER REFERENCE: 014024/0280733
; CURRENT APPLICATION NUMBER: US/09/993, 966
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 60/252, 884
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/291, 109
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/325, 571
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1786
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-993-966-1

```

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	470.00	1786	470	0	0	0
Percent Similarity:	100.00%					
Best Local Similarity:	100.00%					
Query Match:	100.00%					
DB:	10					

US-09-993-966-7 (1-470) x US-09-993-966-1 (1-1786)

```

QY 1 MetGlyLeuHisSerLysProAlaValCysLysArgArgGlySerProGly 20

```

```

Db      159 ATGGGAAATTCATCTCAAGCCGCGCTGTGCAAGCGCAGAGAGCCCGAAAGT 218
Qy      21  AspSerPheAlaValSerAlaIaIaTPAlaArglyIleGluGluTPIleGlyArg 40
Db      219 GACAGCTTCCCGCTGACGCTGCTGGGCTCGAAGGCGATCGAGAGTGAATCGGGAGA 278
Qy      41  GluArgCySPProGlyValSerGlyProArgGluLeuAgluAgluYThIleGly 60
Db      279 CACGCTGCCCGGGCGGTGTCTCGGAGCCCGACAGCTGCGGTGGCGGCACATAGGC 338
Qy      61  ArgSerThrArgGluLeuValIGlyAspValLeuArgSPThrLeuSerGluGluGlu 80
Db      339 CGAAGCACCCGGAGCTCTGTGGCGACGTGTGGAGACAGCTTCAAGCGAAGAGAG 398
Qy      81  AspArgPheArgLeuGluValAlaLeuProProGluIubSPThrApgIlyLeuGly 100
Db      399 GACGACTTTCGGCTGGAAGTGGCCCTGCTCTGAGAAAGACTGACGGGCTGGCAGCG 458
Qy      101 AspGluIubSPMetGluArgValISerGluProCySPProGlySPThrIubSPGlu 120
Db      459 GATGAGAGAGAGATGAGAGAGTGAAGGAACTTGGCCAGGCTTCAAGAACACTGAAG 518
Qy      121 PheGluIubLeuGluGlyAspValISerMetGluIubSPSerArgGluGluTPIle 140
Db      519 TTGGAAGAGCTCCAGTGGAGCTGTCCATGAGAGAGACCGGCGAGAGTGAACCTTC 578
Qy      141 ThrLeuTyRAspPheAspAsnAsnGlyIubValIThrArgIubApgIleThSerLeu 160
Db      579 ACCCTGATGACTTGTGACACAAACGCGAAGGTGACCCGAGAGAGACATCAGCACT 638
Qy      161 HisThrIleTyRGIuValIValAspSerSerValAsnHisSPProThrSerSer 180
Db      639 CACACCATCTATAGAGGTGAGTGTCTCTCAACCACTCCCAACATCCAGAAATG 698
Qy      181 LeuArgValIubSPLeuThrValAlaProAspGlySerGluSerIubSPArgSer 200
Db      699 CTGGGGTAAAGCTCACCTGTGGCCCCCGATGGCAGCCAGAGAAAGAGGCTTCTG 758
Qy      201 AsnGluAlaAspLeuGluSerAlaArgProArgIaGluThIubSPProThrGlu 220
Db      759 AATCAGGCTGACGTCAGAGAGCGAAGGCCCGGACAGAGCAAGCCCATGAGAG 818
Qy      221 ArgSerTrpGluIubSPGlyArgIaProLeuArgPheGluGlyAspSerArg 240
Db      819 CGAGAGCTGGAGAGAGAGAGCGAGCCCGCTCGGTCCAGGCTGACAGCCGCTG 878
Qy      241 GluSerGlyCySPTrHisHisCysValAspGluAsnIleGluArgAsnHisTyR 260
Db      879 CAGCTGGCTGCTCCACCATTTGCTGATGAGAAACATCGAGAGAGAAACCATCT 938
Qy      261 AspLeuAlaGlyIleGluAsnTyRThSerGluPheGlyProGlySerProSer 280
Db      939 GATCTGCCCGGAGAGAAACTACAGCTCCCAATTTGGGCTGCTCCCTTCCGTG 998
Qy      281 GluIubSPSerGluLeuProProArgThSerAspProThrArgSerArgSer 300
Db      999 CAGAGAGCAAGACGCCCCCGCACCTCCATCCATCCATCTCTCCCAAGAAC 1058
Qy      301 GluAlaIleHisIleProHisArgIubSPGluGlyValAspProAlaSerPhe 320
Db      1059 GAAAGCATTCACATCCACACCCGAAAGCCCGAGGCTGAGCCGCGCTCTTCACT 1118
Qy      321 LeuAspThrProIleAlaIubSPValISerGluLeuGluIubSPGluArgI 340
Db      1119 CTGACACCCCAATCGCAAGGTCTCAGAGCTCAGACACGGCTCGGGGAGCAC 1178
Qy      341 GlySerIubSPMetValArgSerProIubSPGluGlyValISerValIleGly 360
Db      1179 GGGAGCAGACATTTGTGAGGTCTCCCAAGGCCCAAGGAGAGTGTGGGTGG 1238
Qy      361 ValAlaArgGlyAlaArgAsnIubSPProLeuGlyProAlaIleProAlaVal 380
Db      1239 GTGGCCAGAGGGGCAAGAAACAGCCCTCTGGGACCCGCACTCTCGGTGTCCC 1298

```

```

Qy      381 SerAlaHisLeuAlaIaSerProAlaLeuProSerLeuAlaProLeuGlyHis 400
Db      1299 TCGGCCACCTGGGTGTCAGCCCGGCTCTCTCCCTCCCTAGCCCCCTCGGCA 1358
Qy      401 LysHisIubSPArgAlaIubSPGluSerGluGluGlyCySPArgGlyLeuGlu 420
Db      1359 AAGCACAAGACCGAGCGCAAGAGAGCCAGAGGAGCTTGGGGGCTTGCAGCA 1418
Qy      421 AlaSerGlyIubSPValIleuGlyArgGluHisLeuArgGluLeuProAlaVal 440
Db      1419 GCCTCAGAGTGGCTCTGTCTGGGGCGGAGACCTTGGGAGGCTGCCGCTGTG 1478
Qy      441 TyRGIubSPGluAlaGlyIubSPProValGluArgHisGluHisIubSP 460
Db      1479 TATGAGAGCAGCGCGGCGAGCGGTCCAGAGACATGACACCAACCATGA 1538
Qy      461 HisHisIubSPTrHisHisPheTyRGIubThr 470
Db      1539 CACCACTTACACACATCTTACAGACA 1568

RESULT 3
US-09-993-966-5
; Sequence 5, Application US/09993966
; Publication No. US20030186232A1
; GENERAL INFORMATION:
; APPLICANT: ROHAN, MICHAEL
; TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NKD PROTEIN,
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF
; FILE REFERENCE: 014024/0280733
; CURRENT APPLICATION NUMBER: US/09/993,966
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 60/252,884
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/291,109
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/325,571
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-993-966-5

Alignment Scores:
Pred. No.: 0          Length: 1859
Score: 470.00        Matches: 470
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match: 100.00%             Indels: 0
DB: 10                      Gaps: 0

US-09-993-966-7 (1-470) x US-09-993-966-5 (1-1859)
Qy      1  MetGlyIubSPLeuHisSerIubSPProAlaIaValCysIubSPArgIubSPProGluGly 20
Db      224 ATGGGAAATTCATCTCAAGCCGCGCTGTGCAAGCGCAGAGAGCCCGAAAGT 283
Qy      21  AspSerPheAlaValSerAlaIaIaTPAlaArglyIleGluGluTPIleGlyArg 40
Db      284 GACAGCTTCCCGCTGACGCTGCTGGGCTCGAAGGCGATCGAGAGTGAATCGGGAGA 343
Qy      41  GluArgCySPProGlyValISerGlyProArgGluLeuArgPheIubSPGluGly 60
Db      344 CACGCTGCCCGGGCGGTGTCTCGAGACCCCGACAGCTGTGGCGGCACCATAGGC 403
Qy      61  ArgSerThrArgGluLeuValIGlyAspValLeuArgSPThrLeuSerGluGluGlu 80
Db      404 CGAAGCACCCGGAGCTCTGTGGCGAGCTGTGAGAGACAGCTTCAAGCGAGAGAGAG 463
Qy      81  AspArgPheArgLeuGluValAlaLeuProProGluIubSPThrApgIlyLeuGly 100

```

Db	464	GAAGCTTTGGCTGGAACTGGCCCTGCTCTGGAGAAAGCTGAAGGGCTGGGACGGGA	523
Qy	101	AspGluLysMetGluArgValSerGluProCysProGlySerLysIleuLys	120
Db	524	GATGAGAAAGAAATGGAGAGATGAGCGAACCTTGGCCAGGCTCCAAAGAAAGACGTGAAG	583
Qy	121	PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGlnGluThrPhe	140
Db	584	TTTGAAGAGCTTCAGTGGCGACCTGTCTCAATGGAGAGAGAACGCCGCGAGAGTGGACTTC	643
Qy	141	ThrIleuTyraerPheAspAsnAsnGluLysValThrArgGluAspIlePheSerIleuLeu	160
Db	644	ACCCGTATGACTTTGACAAACGGCAAGGTCACCCGAGAGACACTCAACAGCTTGCTG	703
Qy	161	HisThrIleTyGluValValAspSerSerValAsnHisSerProThrSerSerLysMet	180
Db	704	CACACCATCTATGAGGTGGTGGACTCTCTGTCTACACCATCCCCAAATCCAGCAAGATG	763
Qy	181	LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysArgSerValLeuVal	200
Db	764	CTGGCGGGTAAACCTCAACCTGGCCCCCGATGGCAGACCAAGCAAGGAGAGCGTCTTGTTC	823
Qy	201	AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeu	220
Db	824	AATCAAGGCTGACCTCGACAGCGCCAAAGGCCCCAGACAGAACCAAGGCCCACTGAGACCTG	883
Qy	221	ArgSerTrpGluLysIleGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu	240
Db	884	CGGAGCTGGAGAAAGAGCAGCGAGGCCGCCGCTCAGGTTCCAGGGTACACAGCCGCCCTGGAG	943
Qy	241	GlnSerGlyCysTyrHisIleCysValAspGluAsnIleGluArgArgAsnHisTyrlleu	260
Db	944	CAGTCTGGCTGTACCACCATTCGTAAGTGAACATGAGAGGGAACCACTCACTTA	1003

```

Db          1544  TATGAGCCAGCGCCGGGACAGCGCTCCAGAGCATATGAGCACCACCAACCATGACAT 1603
Oy          461  HishishistyrHishishaphetYrGlnThr 470
                |||||
Db          1604  CACCACCATTCACACCACTTCTACCAAGACA 1633

RESULT 4
US-09-764-891-5902
; Sequence 5902, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5902
; LENGTH: 2379
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-5902

Alignment Scores:
Pred. No.:      1,3e-185      Length:      2379
Score:          196.00        Matches:      196
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:    41.70%       Indels:       0
Db:             10           Gaps:          0

US-09-993-966-7 (1-470) x US-09-764-891-5902 (1-2379)

```

[illegible]

QY	275	ILYSERPROSERVALAIGLILYSERGILEUPEPROPARQTHSERAPROTHIRAG	294
Db	840	GAGTCCCTCGTGAGCCGAGAGTCAKAACTGCCCCCGCACTTCAATCCACTGCA	899
QY	295	SERATSERNHISGLUPROGLUVALIIEHISILEPROHISATRYLSEPROGLINGLYVALAASP	314
Db	900	TCTGCCTCCATGAGCCGGAAGCCATCCACATCCACACCGAAAGCCCAAGCGGTGCAG	959
QY	315	PROALASERPHENHISPHLEUASPETHRPROILLEALYLVASISERGLULEUGLINGINARG	334
Db	960	CCGGGCTCTTCCACTCTTGTACACCCCAATCGCCAAGTCTCAGAGCTTCAGCAACGG	101
QY	335	LEUATRGILYTHNGINASPGLYSERYSHISPHLEVALARGSERPROLYSALINGILYLS	354
Db	1020	CTCCGGGGGCACCGAGACCGAGACCACTTTGTAGAGTCTCCCCCAAGGCCCAAGGCAAG	107
QY	355	SERVALGILYALGILYHISVALIALARGGLYALARGASMLSEPROLEUGLYPROALA	374
Db	1080	AGTGTGGGTGTGGGCGACGTGGCCAGAGGGGCAAMAACCCCTCTGGAGACCCGGCC	113
QY	375	ILEPROALAVALSERPROSERALAHISLEUAAIASERPROALALEUPEPROSERLEU	394
Db	1140	ATCCTCGGAGGTCCCCCTCGGCCACTGGCTGCAGCCCGGACCTCTCCCCCTCCTA	119
QY	395	ALAPROLEUGILYHISLYLYSHISLYSHISARGALALYSGUSERGINGINLYCYARQ	414
Db	1200	GCCCCCTCGGGCAACAGAGCAACAGCACCGAGCCAGAGAGCAAGCAGGAGGTGCCGG	125
QY	415	GLYLEUGINALAPROLEUHALASERGLYLYPROVALLEUGLYARGGLUHISLEUARGILU	434
Db	1260	GAGCTGACGGACCACTGAGCTCAGGTGACCTGTCTCGGGCGGAGACCACTGGGGAGG	131
QY	435	LEUPROALALEUVALIYTRGLUSERGINALAGLYGINPROVALGINARGHISGLUHIS	454
Db	1320	CTGCCCCGCTTGGTGTGTATGAGACCAAGCCGGGCAACCGGTCCAGAGACATGAGCAC	137
QY	455	HISHSISHISISGLUHISHSISHISISTYRHSISHISPHETRYGLINTHR	470


```

Db      1380 CACCAACCAACATGAATCAATCACCACATTCACCAACTTCTACAGACA 14427

RESULT 5
US-09-764-891-5904
/ Sequence 5904, Application US/09764891
/ Publication No. US20030077808A1
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
/ FILE REFERENCE: PC006
/ CURRENT APPLICATION NUMBER: US/09/764, 891
/ CURRENT FILING DATE: 2001-01-17
/ Prior application data removed - consult PAMM or file wrapper
/ NUMBER OF SEQ ID NOS: 10231
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 5904
/ LENGTH: 2379
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-764-891-5904

Alignment Scores:
Pred. No.: 1,3e-185          Length: 2379
Score: 196.00              Matches: 196
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 41.70%          Indels: 0
DB: 10                      Gaps: 0

US-09-993-966-7 (1-470) x US-09-764-891-5904 (1-2379)

Qy      275 GlycerProSerValAlAGlnYSergIuLeuProProArGThSerAnProThArg 294
Db      840 GGCTCTCCCAAGACCGGAAGCCATCCACATCCACACCGAAAGCCCAAGCGGTGAC 899

Qy      295 SerArgSerHisgIuProGluAlAlHisIlePerHisArgLyProGlnIyValArp 314
Db      900 TCTGGCTCCCAAGACCGGAAGCCATCCACATCCACACCGAAAGCCCAAGCGGTGAC 959

Qy      315 ProAlaSerPheHisPheLeuAspThrProIleAlaIyValSerGluLeuGlnIyArp 334
Db      960 CCGGCTCTCTTCCATTCTCTTGACACCCCAATCGCCAGAGTTCAGAGTCCAGCAAG 1011

Qy      335 LeuArgGlyThrGlnAspGlySerIySerHisPheValArgSerProIyValGlnIyArp 354
Db      1020 CTCCGGGGCAACCCAGACGAGGAGCAAGCACTTGAGAGTCCCCCAAGGCCAGGCAAG 1077

Qy      355 SerValGlyValAlGlyHisValAlaIArgGlyAlaArgAsnIySerProLeuGlyProAla 374
Db      1080 AGTGAGGTGTGGGGCAAGTGGCCAGAGGGGCAAAACACAGCCCTCTGGAGCCCGC 1133

Qy      375 IleProAlaValSerProSerAlaHisIleuAlaIaSerProAlaIleuLeuProSerIeu 394
Db      1140 ATCCCTGGAGGTGTCCTCCCGCCCACTGGCTGCGAGCCGCGCCCTCTCTCCCTCTCA 1199

Qy      395 AlaProLeuGlyHisIyLeuIySerHisIyArgAlaIyGluSerGlnGlnIyCyArg 414
Db      1200 GCGCCCTCGGGCAACAGAGCAAGCAACGACCGAGCAAGAGGCAAGAGGAGGCTGCCG 1255

Qy      415 GlyLeuGlnAlaProLeuAlaSerGlyIyProValIleuGlyArGluHisIleuArgIu 434
Db      1260 GGCTCTGAGGCAACCACTGGCTCAAGTGGCTCTGTCGGGCGGAGAGCACTGGGGAG 1311

Qy      435 LeuProAlaIeuValIyTrGlnSerGlnAlaGlyIyProValIleuGlyArGluHisIleuArgIu 454
Db      1320 CTGCGCGCTGTGGTGTGTATGAGAGCAAGGCGCGGGAGCCGGTCCAGAGATAGAGC 1375

Qy      455 HisHisHisHisIyGluHisHisHisHisIyTrHisHisPheTyGlnThr 470
Db      1380 CACCAACCAACATGAATCAATCACCACATTCACCAACTTCTACAGACA 14427

RESULT 6
US-10-087-192-178

```

Sequence 179, Application US/10087192					
Publication No. US2002018256A1					
GENERAL INFORMATION:					
APPLICANT: Morris, David W.					
APPLICANT: Engelhard, Eric K.					
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR					
TITLE OF INVENTION: CANCER					
FILE REFERENCE: 52945200122					
CURRENT APPLICATION NUMBER: US/10/087,192					
CURRENT FILING DATE: 2002-03-01					
PRIOR APPLICATION NUMBER: US 09/747,377					
PRIOR FILING DATE: 2000-12-22					
PRIOR APPLICATION NUMBER: US 09/798,586					
PRIOR FILING DATE: 2001-03-02					
NUMBER OF SEQ ID NOS: 2059					
SOFTWARE: FastSeq for Windows Version 4.0					
SEQ ID NO 178					
LENGTH: 106315					
TYPE: DNA					
ORGANISM: Homo sapiens					
FEATURE:					
NAME/KEY: misc_feature					
LOCATION: (1)...(106315)					
OTHER INFORMATION: n = A,T,C or G					
US-10-087-192-178					
Alignment Scores:					
Pred. No.:	3,29e-184	Length:	106315		
Score:	196.00	Matches:	196		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	41.70%	Indels:	0		
DB:	13	Gaps:	0		
US-09-993-966-7 (1-470) x US-10-087-192-178 (1-106315)					
QY	275	GIYSErProSeRValAlAGInLYSeRGInLeuPProPrOArGThSeRAnPProThrArG	294		
DB	95728	GGCTCCCTCCCTCCGGGCGCCAGAAAGTCAGAACTCCCCCGGCACTCCACCTCCACTCCAA	95787		
QY	295	SeRarSeRnHAGInPProGInWalAlEnHsAlleProHIsARghbPProGInGInYValaAP	314		
DB	95788	TCTGGCTCCCAATGAGCGGGAAGGCATCCATCCACACCGAAAGGCCCAAGGCGTGAC	95847		
QY	315	ProLaSeRPhenHsPheLeuASeRThrProLleAlaYSeRValaSeRGInGInGInArG	334		
DB	95848	CCGGCTCTCTTCCACTTCTCTTGACACCCCAATGCGCAAGTCTCAAGCTCCAGCAACGG	95907		
QY	335	LeuARGLYThrGInaPbGlySeRtLYSeRtLYbPheValaLYSeRProLYaLaGInGInLYbS	354		
DB	95908	CTCCGGGGGCAACCCAGGACGGAGCAAGCACTTGTGAGGTCGCCCAAGGCCCGAGGCAAG	95967		
QY	355	SeRValGInYValGInHsValAlaAaRGInYAlaArgAaInLYbProProLeuGInYProAla	374		
DB	95968	AGTGTGGGTGTGGGCCACGTGGCAAGGGGCAAGAAAGAGCCCTCTGGAGACCCGGC	96027		
QY	375	IlleProAlaValaSeRProSeRAlaHIsbLeuAlaAlaSeRProAlaLeuLeuPProSeRLeu	394		
DB	96028	ATCCCTCGGGTGTCCCTCCCTCCGCCCACTGGCTGCGCAAGCCGGCCCTCTCCCTCCCTTA	96087		
QY	395	AlaProLeuGInYbLYbLYbHIsbLYbHIsbArgAlaLYbGInSeRGInGInGInCYaArG	414		
DB	96088	GCCCCCTCCGGGCAAGAGCAAGCAAGCAACCGAGCCAGAGAGCCAGAGGGGCTCCGG	96147		
QY	415	GIYLeuGInAlaProLeuAlaSeRGInYbLYbProValLeuGInYArgGInHIsbLeuArgGIn	434		
DB	96148	GGCTTGAGGACACCACTGGGCTAGGTGGCCCTGTCTGGGCGGGAGACCTGGGGAG	96207		
QY	435	LeuProAlaLeuValaLYbLYbSeRGInAlaGInGInProValGInaRghHsGInHIs	454		
DB	96208	CTGCCCGCCCTTGGTGGTGTATGAGAGCCAGGCGGGGCAACGGGTCCAGAGACTAGACAC	96267		
QY	455	HIsHIsHIsHIsGInHIsHIsHIsHIsHIsHIsHIsHIsHIsHIsHIsHIsHIsHIsHIsHIsHIs	470		

Db 96268 CACCACCACCATGAACATCACACCATTCACACCATTTCTACCGAGACA 96315
RESULT 7
US-09-993-966-21
Sequence 21, Application US/09993966
Publication No. US20030186232A1
GENERAL INFORMATION:
APPLICANT: ROHMAN, MICHAEL
TITLE OF INVENTION: HUMAN AND NON-HUMAN PRIMATE HOMOLOGUES OF NPD PROTEIN,
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING, AND USES THEREOF
FILE REFERENCE: 014024/0280733
CURRENT APPLICATION NUMBER: US/09/993,966
CURRENT FILING DATE: 2001-11-27
PRIOR APPLICATION NUMBER: 60/252,864
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/291,109
PRIOR FILING DATE: 2001-05-16
PRIOR APPLICATION NUMBER: 60/325,571
PRIOR FILING DATE: 2001-10-01
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 21
LENGTH: 590
TYPE: DNA
ORGANISM: Homo sapiens
US-09-993-966-21
Alignment Scores:
Pred. No.: 3,97e-185 Length: 590
Score: 195.00 Matches: 195
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 41.49% Indels: 0
Gaps: 0
US-09-993-966-7 (1-470) x US-09-993-966-21 (1-590)
QY 276 SerProSerValAlaGlnLysSerGluLeuProArgThrGlnSerProThrArgSer 295
Db 3 TCCCTCTCCGTGGCCCGAGAGTCAGAACTGCCCCCGGACCTTCATCCCATCTCGATCT 62
QY 296 ArgSerHisGluProGluAlaIleHisIleProHisArgLysProGlnIlyValAspPro 315
Db 63 CGCTCCCATGAGCGGAGGACCATCATCCCAACCGAAAGCCCAAGGGGTGACCGG 122
QY 316 AlaSerPheHisPheLeuAspThrProIleAlaLysValSerGluLeuGlnIlyArgLeu 335
Db 123 GCCCTCTTCCACTTCTTGACACCCCATCGCCCAAGGCTTCAGAGCTCCAGCAACGGCTC 182
QY 336 ArgGlyThrGlnAspGlySerLysHisPheValArgSerProLysValGlnIlyLysSer 355
Db 183 CGGGGACCCGAGCGGAGGACCACTTGTGTGAGTCCCCCAAGGCCCAAGGCAAGAGT 242
QY 356 ValGlyValGlnHisValAlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIle 375
Db 243 GTGGGTGTGGGCGCACTGTGGCGAGGGGCAAGAAACAAGCCCTCTGGGACCGGCATC 302
QY 376 ProAlaValSerProSerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAla 395
Db 303 CCTTCGGGTGTCCCTCCCGCCACCTGCTCCAGCCCGGCTCTCCCTCCCTCCCTAGCC 362
QY 396 ProLeuGlyHisLysLysHisLysHisLysHisLysHisLysHisLysHisLysHis 415
Db 363 CCCCCTGGGCGACAGAGACACAGACACCGAGCCCAAGGACCGAGGCTGGCGGAGGC 422
QY 416 LeuGlnAlaProLeuAlaSerGlyIlyProValLeuGlyArgGlyLysHisLysArgGly 435
Db 423 CTGCAAGGACACCACTGGCTCAGTGGCTCTCTCTGGGGCGGAGCACTGCGGAGGTG 482
QY 436 ProAlaLeuValValIlyArgLysSerGlnAlaGlyLysProValGlnAlaGlnHis 455
Db 483 CCGGCTTGT 542

QY 456 HisHisHisGluHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 470
Db 543 CACCACCATGAACATCACACCATTCACACCATTTCTACCGAGACA 587
RESULT 8
US-09-918-995-8686
Sequence 8686, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 8686
LENGTH: 427
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(427)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-8686
Alignment Scores:
Pred. No.: 2,92e-128 Length: 427
Score: 138.00 Matches: 138
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 29.36% Indels: 0
Gaps: 0
US-09-993-966-7 (1-470) x US-09-918-995-8686 (1-427)
QY 285 LeuProProArgThrSerArgProThrArgSerArgSerHisGluProGluAlaIleHis 304
Db 4 CTCGCCCCCGGACCTTCATCCCATCTCGATCTCGCTCCCATGAGCGGAAAGCCATTCAC 63
QY 305 IleProHisArgLysProGlnIlyValAspProAlaSerPheHisPheLeuAspThrPro 324
Db 64 ATCCCAACCGAAAGCCCGCAAGGCGTGGACCGGCTCTCTTCCATCTCTTGACACCCCA 123
QY 325 IleAlaLysValSerGluLeuGlnIlyArgLeuArgGlyThrGlnAspGlySerLysHis 344
Db 124 ATCGCCAAAGGTCTCAGAGCTCCAGCAACGGCTCCGGGGCACCCAGGACGGGACCAAGCAC 183
QY 345 PheValArgSerProLysValGlnIlyLysSerValGlyValGlnHisValAlaArgGly 364
Db 184 TTTGTGAGTCTCCCGCAAGGCCAGGCGACAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 243
QY 365 AlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerProSerAlaHisLeu 384
Db 244 GCAGAAACAAGGCCCTCTGGGACCGGCATCTCTCGGTGTCTCCCTCCGCGCACCTG 303
QY 385 AlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLysLysHis 404
Db 304 GCTGCGAGCGCGGCTCTCTCCCTCCCTGCGGCTCTCCGCGGACCAAGAGCACAAGCAC 363
QY 405 ArgAlaLysGluSerGlnIlyGlyArgGlyLysHisLysHisLysHisLysHisLysHis 422
Db 364 CGAGCCAAAGAGGACCAACAGGGCTGCGGGGCTCTGCACACCACTAGGCTCA 417
RESULT 9
US-09-993-966-18
Sequence 18, Application US/09993966
Publication No. US20030186232A1
GENERAL INFORMATION:


```

; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Mus sp.
US-09-993-966-2

Alignment Scores:
Pred. No.: 7.69e-39 Length: 1416
Score: 49.00 Matches: 49
Percent Similarity: 100.00% Conservaive: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.43% Indels: 0
DB: 10 Gaps: 0

US-09-993-966-7 (1-470) x US-09-993-966-2 (1-1416)

Qy 131 GluGluAspSerArgGlnGluTrpThrPheThrLeuTyrAspPheAspAsnGlyLys 150
Db 391 GAGGAGGACGCGCGGAGAGTGACTTTCACTATATAGACTTCGACAAATGGGAAA 450
Qy 151 ValThrArgGluAspIleThrSerLeuLeuHisThrIleTyrGluValValAspSerSer 170
Db 451 GTGACCCGTGAGGACATTCACAGCTTGTCATACCATCTATGAAAGTGTTGACTCTCT 510

Qy 171 ValAsnHisSerProThrSerSerLys 179
Db 511 GTGAACCATTCCTCCCATCATCAAGCAAG 537

RESULT 13
US-10-087-192-176
; Sequence 176, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 52945200122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 176
; LENGTH: 2807
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-087-192-176

Alignment Scores:
Pred. No.: 1.38e-38 Length: 2807
Score: 49.00 Matches: 49
Percent Similarity: 100.00% Conservaive: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.43% Indels: 0
DB: 13 Gaps: 0

US-09-993-966-7 (1-470) x US-10-087-192-176 (1-2807)

Qy 131 GluGluAspSerArgGlnGluTrpThrPheThrLeuTyrAspPheAspAsnGlyLys 150
Db 333 GAGGAGGACGCGCGGAGAGTGACTTTCACTATATAGACTTCGACAAATGGGAAA 392
Qy 151 ValThrArgGluAspIleThrSerLeuLeuHisThrIleTyrGluValValAspSerSer 170
Db 393 GTGACCCGTGAGGACATTCACAGCTTGTCATACCATCTATGAAAGTGTTGACTCTCT 452
```

```

Qy 171 ValAsnHisSerProThrSerSerLys 179
Db 453 GTGAACCATTCCTCCCATCATCAAGCAAG 479

RESULT 14
US-10-016-157A-126
; Sequence 126, Application US/10016157A
; Publication No. US20020192220A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Ghosh, Malavika
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and Prot
; FILE REFERENCE: DEX-0253
; CURRENT APPLICATION NUMBER: US/10/016,157A
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,717
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 250
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 126
; LENGTH: 1950
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-016-157A-126

Alignment Scores:
Pred. No.: 1.01e-35 Length: 1950
Score: 46.00 Matches: 46
Percent Similarity: 100.00% Conservaive: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.79% Indels: 0
DB: 13 Gaps: 0

US-09-993-966-7 (1-470) x US-10-016-157A-126 (1-1950)

Qy 20 GlYAspSerPheAlaValSerAlaIleTrrPAlaArgLysGlyIleGlnGluTrpIleGly 39
Db 838 GTGACGACTTCCTCCGTGACGCTGCTGCGCTCGAAGGCATTCAGAGTGATCGG 897
Qy 40 ArgGlnArgCybProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIle 59
Db 898 AGACAGGCGTCCCGGGCGGTGTCTCGGAGCCCGACAGCTGCGTTGGCGGACACATA 957
Qy 60 GlYArgSerThrArgGlu 65
Db 958 GGCCGAGCACCCGGGAA 975

RESULT 15
US-10-029-386-7990
; Sequence 7990, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GI
; FILE REFERENCE: ABOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7990
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Homo sapiens
FEATURE:
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4
```

```

; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8
; OTHER INFORMATION: SWISSPROT HIT: Q23977, EVALUE 7.70e-02
; OTHER INFORMATION: NT HIT: g14916432, EVALUE 9.00e-71
; OTHER INFORMATION: EST_HUMAN HIT: BF920913.1, EVALUE 3.00e-68
US-10-029-386-7990

```

```

Alignment Scores:
Pred. No.:      3.69e-35      Length:      598
Score:          45.00         Matches:      45
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0
Query Match:     9.57%       Indels:          0
DB:              15          Gaps:            0

```

US-09-993-966-7 (1-470) x US-10-029-386-7990 (1-598)

```

QY      20  G|YAspSerPheAlaValSerAlaAlaTrpAlaArgLyegLytleGluGluTrpIleGly 39
      |||
Db      229  GGTGACAGCTTCGCGGTGAGGCTGCTGCGGCTCGAAGGCGCATCGAGAGTGATCGCG 288
      |||
QY      40  ArgGlnArgCysProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIle 59
      |||
Db      289  AGACAGCGCTGCGCGCGGTGTCTCGGAGACCCCGACAGCTGCGGTGCGGCGCACCA 348
      |||
QY      60  GlyArgSerThrArg 64
      |||
Db      349  GCCCGAAGCAACCCGG 363
      |||

```

Search completed: December 30, 2004, 14:01:27
 Job time : 995 secs

THIS PAGE IS BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 30, 2004, 04:02:40 ; Search time 848 Seconds
(without alignments)
2909.467 Million cell updates/sec

Title: US-09-993-966-7
Perfect score: 470
Sequence: 1 MGKLSKPAVCKRKRESPEC.....RHEHHHHHHHHHHFYQT 470

Scoring table: **QWGO!**
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 413486 seqs, 2624710521 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8266295

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xlp
-Q/cgm2.1/USPTO.spool.p/US09993966/runat.29122004.130828.9372/app.query.fasta_1.647
-DB=N Geneseq.23Sep04 -QFMT=fasta -SUFFIX=01g.rng -MINMATCH=0.1 -LOOFCU=0
-LOOPEXT=0 -INITs=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=humand0.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pico -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09993966 @CGN 1.1 885 @runat.29122004.130828.9372 -NCPU=6 -ICPU=3
-NO MMAP -LARGEJQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N_Geneseq.23Sep04.*

1: geneseqn1980s.*
2: geneseqn1990s.*
3: geneseqn2000s.*
4: geneseqn2001as.*
5: geneseqn2001bs.*
6: geneseqn2002as.*
7: geneseqn2002bs.*
8: geneseqn2003as.*
9: geneseqn2003bs.*
10: geneseqn2003cs.*
11: geneseqn2003ds.*
12: geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	470	100.0	1859	10	AAL51511
2	383	81.5	1438	3	AAAG63925
3	196	41.7	2379	4	AAL03214
4	196	41.7	2379	4	AAL03216
5	195	41.5	590	10	AAL51525
6	138	29.4	427	9	ACH21474

7	107	22.8	1561	5	AA92082	AA92082 DNA encod
8	72	15.3	1743	5	AA378752	AA378752 DNA encod
9	72	15.3	1743	5	AA378752	AA378752 DNA encod
10	72	15.3	1743	5	AA378752	AA378752 DNA encod
11	49	10.4	148	10	AAL51522	AAL51522 Mouse Nkd
12	49	10.4	1401	10	AAL51512	AAL51512 Mouse Nkd
13	49	10.4	1416	4	AAH28343	AAH28343 Nucleotid
14	46	9.8	1950	6	ABQ73820	ABQ73820 Human col
15	45	9.6	358	12	ACH74795	ACH74795 Human gen
16	45	9.6	1731	3	AAAG3924	AAAG3924 DNA encod
17	44	9.4	133	12	ACH88495	ACH88495 Human gen
18	44	9.4	134	10	AAL51518	AAL51518 Human Nkd
19	42	8.9	138	10	AAL51524	AAL51524 Human Nkd
20	36	7.7	639	8	ACD05637	ACD05637 cDNA enco
21	35	7.4	107	10	AAL51520	AAL51520 Human Nkd
22	32	6.8	96	10	AAL51521	AAL51521 Human Nkd
23	29	6.2	179	12	ACH74420	ACH74420 Human gen
24	28	6.0	179	12	ACH88120	ACH88120 Human gen
25	27	5.7	85	10	AAL51523	AAL51523 Human Nkd
26	22	4.7	67	10	AAL51519	AAL51519 Human Nkd
27	27	3.6	1285	3	AAAG3926	AAAG3926 DNA encod
28	28	3.6	1307	3	AAAG3927	AAAG3927 DNA encod
29	17	3.6	1817	6	ABK34945	ABK34945 Human CDN
30	17	3.6	1885	10	ADP82455	ADP82455 Leukaemia
31	17	3.6	1959	12	ADO22400	ADO22400 Human sof
32	17	3.6	1966	4	AA327117	AA327117 cDNA enco
33	17	3.6	1966	4	ADP93285	ADP93285 Human CDN
34	14	3.0	551	10	ADP81081	ADP81081 Leukaemia
35	13	2.8	2723	4	AAK75944	AAK75944 Human imm
36	11	2.3	34	10	AAL51517	AAL51517 Human Nkd
37	10	2.1	574	4	AAL10301	AAL10301 Probe #23
38	10	2.1	574	4	ABAS1944	ABAS1944 Human foe
39	10	2.1	574	4	AAL11552	AAL11552 Probe #23
40	10	2.1	574	4	ABA21762	ABA21762 Probe #22
41	10	2.1	574	4	AAK25674	AAK25674 Human bon
42	10	2.1	574	4	AAK00231	AAK00231 Human bra
43	10	2.1	574	4	AB525257	AB525257 Human liv
44	10	2.1	574	5	AAL00237	AAL00237 Probe #22
45	10	2.1	574	6	AB500249	AB500249 Human gen

ALIGNMENTS

RESULT 1
ID AAL51511 standard; DNA, 1859 BP.

XX	AC	AAL51511;	
XX	DT	24-APR-2003	(first entry)
XX	DE	Human Nkd (hNkd)	protein coding sequence.
XX	KW	Human, gene; ds; gene therapy; Nkd; hNkd; hnt signalling pathway; cancer;	
XX	KW	colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer.	
XX	OS	Homo sapiens.	
XX	FT	Key	Location/Qualifiers
XX	FT	CDS	224..1636
XX	FT		/tag= a
XX	FT		/product= "Human Nkd protein"
XX	PD	21-NOV-2002.	
XX	PF	27-NOV-2001; 2001MO-US044092.	
XX	PR	27-NOV-2000; 2000US-0252884P.	
XX	PR	16-MAY-2001; 2001US-0291109P.	
XX	PR	01-OCT-2001; 2001US-0325571P.	

PA (CHIR) CHIRON CORP.
 XX
 PI Rohan M, Chan V, Yan D;
 XX WPI, 2003-129303/12.
 DR P-PSDB; AAO16389.
 XX
 PT New human and non-human primate homologues of Nkd protein, and Nkd genes,
 PT useful for treating cancer involving aberrant Wnt signaling, e.g. colon
 PT cancer, head and neck cancer, ovarian cancer, or breast cancer.
 XX
 PS Claim 4; Fig 3; 99p; English.
 XX
 CC The invention comprises the amino acid and coding sequence of the human
 CC Nkd protein (hNkd) - a regulator of Wnt signalling pathways. The hNkd DNA
 CC and protein sequences are useful for the treatment of cancer involving
 CC aberrant Wnt signalling (e.g. colon cancer, head and neck cancer, ovarian
 CC cancer and breast cancer). The present DNA sequence encodes the human Nkd
 CC protein of the invention
 XX
 SQ Sequence 1859 BP; 434 A; 587 C; 551 G; 287 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0 Length: 1859
 Score: 470.00 Matches: 470
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-09-993-966-7 (1-470) x AAL51511 (1-1859)

QY 1 MetGlyysLeuHisSerlySPROAlaValAlcySylsArGArGlyUserProGly 20
 Db 224 ATGGGGAAACTTCACTCCAAAGCCGCGCGCTGTGCAAGCGCAGAGAGCCGGAAGGT 283
 QY 21 AspSerPheAlaValSerAlaAlaTPrAlaArglySgLytleGluGlyTPrileGlyArg 40
 Db 284 GACAGCTTCGCGCGTGAAGCGCTGCTGGAGTGGAGGAGCATCGAGAGTGGATGGGAGA 343
 QY 41 GlnArgCySPProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrileGly 60
 Db 344 CAGCGCTGCCCGGGCGGTGTCTCGGAGCCCGACAGCTGGCGTGGCGGACCATAGGC 403
 QY 61 ArgSerThrArgGlyLeuValAlaGlyAspValLeuArgPheThrleuSerGlyGluGly 80
 Db 404 CGAAGCACCCGGAGGCTGTGGGAGACGTGTGAAGACACGCTCAACGAGGAAGAG 463
 QY 81 AspAspPheArgLeuGlyValAlaLeuProProGlyThrArgPlyLeuGlySerGly 100
 Db 464 GACGACTTTCGGCTGGAGAGTGGCTGCTCTTGAAGACTGACGGGCTGGGACGGGA 523
 QY 101 AspGlylySylsMetGlyArgValSerGlyProCySPProGlySerlySylsGlnLeu 120
 Db 524 GATGAGAAAGAGATGAGAGAGTGAAGGACCTGCCAGGCTCAAGAAAGCACTGAAG 583
 QY 121 PheGluGluLeuGlnCySPAspValSerMetGluGluAspSerArgGlnGlyTPrPhe 140
 Db 584 TTGGAAGAGCTCCAGTGTGATGAGAGAGACAGCCGGGAGGTGAGCTTC 643
 QY 141 ThrleuTyraSPpPheAspAsnGlylyValAlaThrArglyuAspIleThrSerleu 160
 Db 644 ACCCTGATGACTTGAACACACGAGGATCAACCGAGAGGACATCACAGCTGTG 703
 QY 161 HisThrIleTyrglyValAlaAspSerSerValAsnHisSerProThrSerSerlyMet 180
 Db 704 CACGACCTATATGAGGTGTGACTCTCTGTCAACCACTCCCAACATCCAGCAAGATG 763
 QY 181 LeuArgVallySylsLeuThrValAlaProAspGlySerGlnSerlySylsSerValleuVal 200
 Db 764 CTGGGGTAAAGCTCACCGTGGCCCGATGGACGACGAGCAAGAGAGGTCTTGTCT 823
 QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGlyuThrlySPProThGluAspLeu 220

Db 824 AATCAGGCTGACTGCGAGAGCGCAGAGGCCCGGAGAGAGACAAGCCCATGAGGACTG 883
 QY 221 ArgSerTPrGlylySylsGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGly 240
 Db 884 CGAGGCTGGGAGAAAGAGCAGAGCGAGCCCGCTCAGGTTCAGAGGTGACACCGCTGAG 943
 QY 241 GlnSerGlyCySPTyraHisSylsValAspGlyuAsnIleGluArgArgAsnHisTyrl 260
 Db 944 CAGTCTGGCTGCTACCACTTGGCTGATGAGAAACATCGAGAGAGAAACCACTACTTA 1003
 QY 261 AspLeuAlaGlytleGluAsnTyraSerGlnPheGlyProGlySerProSerValAla 280
 Db 1004 GATCTGCCCGGATAGAAACTACAGCTCCCAATTGGAGCTGCCCTTCGTGGGCC 1063
 QY 281 GlnlySPSerGlyLeuProProArgThrSerAspProThrArgSerArgSerlySyl 300
 Db 1064 CAGAAAGTCAGAACTGCCCCCGCACCTCCAAATCCCACTGATCTCGCTCCCATGAGCCG 1123
 QY 301 GluAlaIleHisIleProHisArglySPProGlnGlyValAspProAlaSerPheIlePhe 320
 Db 1124 GAAGCATCCACATCCACACCGAAAGCCCAAGGCGTGAACCGGCTCTCTTCACTTC 1183
 QY 321 LeuAspThrProIleAlaValSerGlyLeuGlnGlnArgLeuArgGlyThrGlnAsp 340
 Db 1184 CTTGACACCCCATTCGCGCAAGTCTCAGAGCTCCAGCAAGGGCTCCGGGACACCAAGAC 1243
 QY 341 GlySerlySPheValArgSerProlySPAlaGlnGlylySPSerValGlyValGlyHis 360
 Db 1244 GGGAGCAAGCACTTGTGAGGTCCCCCAAGGCCCAAGGAGTGTGGGTGTGGGCAC 1303
 QY 361 ValAlaArgGlyAlaArgAsnlySPProProLeuGlyProAlaIleProAlaValSerPro 380
 Db 1304 GTGGCCAGAGGGGCAAGAACAGCCCTCTGGAGCCCGCATCTCGGGGTGTCCCCC 1363
 QY 381 SerAlaHisLeuAlaAspProAlaLeuLeuProSerLeuAlaProLeuGlyHis 400
 Db 1364 TCGGCCACCTGGGTGCAAGCCCGGCTCTCCCTCCCTAGGCCCTCGGGACAAAG 1423
 QY 401 LysHisIleHisValArgAlaValSerGlnGlnGlyCySPArglyLeuGlnAlaProLeu 420
 Db 1424 AAGCACAAAGCAGAGCCAGAGAGCCAGAGGCTGCCGGGCTCGAGGCAACCACTG 1483
 QY 421 AlaSerGlyGlyProValLeuGlyArgGlyuHisLeuArgGlyLeuProAlaLeuValVal 440
 Db 1484 GCTTCAGGTGCTCTGTCTGGGCGGAGCACTGCGGAGCTGCCGCTTGTGTGTG 1543
 QY 441 TyrglySPSerGlnAlaGlyGlnProValGlnArgHisIleGlyHisHisIleGlyHis 460
 Db 1544 TATGAGAGCAGAGCGGCGGAGCGGTCAAGACATGAGCACCAACCAATGAACAT 1603
 QY 461 HisHisIleTyraHisIlePheTyrglnThr 470
 Db 1604 CACCACCATTAACCACTTCTACAGACA 1633

RESULT 2
 AAA63925
 ID AAA63925 standard; DNA; 1438 BP.
 XX
 AC AAA63925;
 XX
 DT 04-DEC-2000 (first entry)
 XX
 DE DNA encoding protein related to Drosophila naked cuticle polypeptide.
 XX
 KW Nkd gene; naked cuticle gene; segment-polarity gene; Wnt signalling;
 XX gene therapy; Nkd defect; cancer; ss.
 XX
 OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 8..1420
 FT /*tag= a

FT /product= "naked cuticle polypeptide"

XX WO200049034-A1.

PN 24-AUG-2000 *a dte*

PD 17-FEB-2000; 2000WO-US004188.

PF 17-FEB-1999; 99US-0120646P.

XX 17-FEB-1999; 99US-0120646P.

XX (STRD) UNIV LEIAND STANFORD JUNIOR.

PI Scott M, Zeng W, Wharton K;

XX WPI, 2000-571967/53.

DR P-PSDB; AAB08216.

XX An isolated nucleic acid molecule useful for analyzing (genetic

PT predisposition to) a disease state and for therapeutic purposes e.g.

PT treatment of cancer comprises a sequence encoding a naked cuticle

PT protein.

XX Claim 3; Page 46-48; 58pp; English.

XX The present sequence encodes a protein related to the Drosophila Nkd

CC (naked cuticle) polypeptide. In Drosophila, Nkd is a segment-polarity

CC gene whose expression is induced by Wnt signalling. The Nkd polypeptide

CC acts to antagonize Wnt signalling. Nkd may link ion fluxes to the

CC regulation of Wnt signal potency, duration or distribution. The Nkd

CC polypeptides can be used for identifying homologous or related

CC proteins, to modulate the expression or function of Nkd polypeptides, and

CC in studying associated physiological pathways. Nkd polynucleotides can

CC also be used in gene therapy to treat disorders associated with Nkd

CC defects. They may also be used for therapeutic purposes e.g. treatment of

CC cancer

XX Sequence 1438 BP; 319 A; 478 C; 439 G; 202 T; 0 U; 0 Other;

SO

Alignment Scores:

Pred. No.:	0	Length:	1438
Score:	383.00	Matches:	469
Percent Similarity:	99.58%	Conservative:	0
Beet Local Similarity:	99.58%	Mismatches:	1
Query Match:	81.49%	Indels:	2
DB:	3	Gaps:	0

US-09-993-966-7 (1-470) x AAA63925 (1-1438)

QY 1 MetGlyLeuSerProAlaValCysLeuArgArgGlyLeuSerProGly 20

DB 8 ATGGGGAACTTCATCTCAAGCCGCGCGTGTGCAAGCGCAGGAGACCCCGAAGGT 67

QY 21 AspSerPheAlaValSerAlaAlaTrpAlaArgLeuGlyIleGluIleGlyArg 40

DB 68 GACAGCTTCCCGCTGACCGCTGCTGAGGAGCATCGAGAGTGGATCGGAGA 127

QY 41 GluArgCysProGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGly 60

DB 128 CACGCGTCCCGCGGCGGTGTCTCGGAGCCCGACAGCTGCGGTGGCGGACCATGAGGC 187

QY 61 ArgSerThrArgGlnLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80

DB 188 CGAAGCAGCCCGGAGCTCGTGGGAGCAGTGTGAGACACGCTCAGCGAGGAGAGAG 247

QY 81 AspAspPheArgLeuGlnValAlaLeuProProGlyLeuSerGlyLeuGlySerGly 100

DB 248 GAGGACTTTCGGCTGGAAGTGGCTCTCTCTAGAGAGACTGAGGCTGGGAGCGGA 307

QY 101 AspGluValLeuMetGluArgValSerGluProCysProGlySerGlyLeuGlu 120

DB 308 GATGAGAAAGATGAGAGAGTGAAGGAGACCTGCGCAGGCTCAAGAGACAGCTGAG 367

QY 121 PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGlnGluTrpThrPhe 140

DB 368 TTGGAAGCTCCAGTCCGAGTGTCTCAAGAGAGACAGCCGCGAGTGAACCTTC 427

QY 141 ThrLeuTyraAspPheAspAsnGlyValThrArgGluAspIleThrSerLeuLeu 160

DB 428 ACCCTGTATCACTTTGACCAACAGCGAAGTACCCGAGAGGACATCACAGCTTGCTG 487

QY 161 HisThrTleArgGluValValAspSerSerValAsnHisSerProThrSerSerLeuMet 180

DB 488 CACACCATCTATAGGGTGGTGACTCTCTGTCAACCACTCCCAATCCAGCAATG 547

QY 181 LeuArgValLeuThrValAlaProAspGlySerGlnSerIleValAspSerValLeuVal 200

DB 548 CTGGGGTAAAGCTCACGTGGCCCGATGGAGCCAGCAAGAGAGAGGCTTGTTC 607

QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrArgProThrArgLeu 220

DB 608 AATCAGGCTGACCTGCGAGAGCGCAGGCCCGAGCAGACAGCAAGCCACTGAGAGCTG 667

QY 221 ArgSerTrpGluValLeuArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240

DB 668 CGGAGCTGGGAGAAAGCAGCGAGCCCGCTCAGGTTCAGAGGTGACAGCCGCTGAG 727

QY 241 GluSerGlyCysTrpArgHisCysValAspGluAsnIleGluArgArgAsnHisArgLeu 260

DB 728 CAGCTGTGCTGTACCACTTCGTTGATGAGAACATCAAGAGAGAAACCTACTTCA 787

QY 261 AspLeuAlaGlyIleGluAsnTrpThrSerGlnPheGlyProGlySerProSerValAla 280

DB 788 GATCTCCCGGGATAGAAACTACACTCCCAATTTGGCTGGCTCCCTCGTGGCC 847

QY 281 GluLeuSerGluLeuProProArgThrSerAsnProThrArgSerArgSerHisGluPro 300

DB 848 CAGAGTCAGAAAGTGGCCCGCCACCTT-CAATCCCACTGATCTGCTCCCAATGAGCC 906

QY 300 OGluAlaIleHisIleProHisArgLeuProGlnGlyValAspProAlaSerPheHisPhe 320

DB 907 GAAAGCATCATCACTCCACACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 966

QY 320 GluAspThrProIleAlaValSerGluLeuGlnGlnArgLeuArgGlyThrGlnAs 340

DB 967 CTTTGAACCCCAATGCCAAGGCTTCAGAGCTCCAGCAAGGCTCGGGGCACTCAAGA 1026

QY 340 PGIYSerLeuHisPheValArgSerProGlyValGlnGlyValSerValGlyValGln 360

DB 1027 CGGAGCAGACACTTGTGTGAGTCCCGCAAGGCCAGGAGGAGTGTGGTGTGGGCCCA 1086

QY 360 ValAlaArgGlyValArgAsnLeuProProLeuGlyProAlaIleProAlaValSerP 380

DB 1087 GTGGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1146

QY 380 SerAlaHisIleValAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHis 400

DB 1147 CTCGCGCACTGGCTGCCAGCCCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1206

QY 400 ValValSerValArgAlaValSerGlnGlnGlyCysArgGlyLeuGlnAlaProLe 420

DB 1207 GAAAGCAGACACCGAGCCAGGAGAGCCAGAGGAGCTGCGGAGGCTGCGAGGAGCAACT 1266

QY 420 ValAspGlyValProValLeuGlnArgGlnHisIleValArgGluLeuProAlaVal 440

DB 1267 GGCTTCAGTGGCTGTCTCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1326

QY 440 TYGLeuSerGlnAlaGlyGlnProValGlnArgHisGluHisHisHisHisGluHis 460

DB 1327 GTATGAGAGCAGAGCGGAGGAGCGGTCCAGAGACATGAGACACACACCATGAGACA 1386

QY 460 HisHisHisHisHisHisPheTrpGlnThr 470

DB 1387 TCACCAACATTAACACACTTCAACAGACA 1417

RESULT 3

AA03214

ID AAL03214 standard; DNA: 2379 BP.
XX AAL03214;
AC
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 5902.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN W0200155320-A2.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US001339.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 25-SEP-2000; 2000US-0235484P.
PR 26-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 27-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244647P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.


```
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234978P.
PR 25-SEP-2000; 2000US-0234988P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240950P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249264P.

PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 06-DEC-2000; 2000US-0251866P.
PR 06-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 03-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen is
XX used in preventing, treating or ameliorating a medical condition.
XX
XX PS Disclosure; SEQ ID NO 5904; 1297bp + Sequence Listing; English.
XX
XX CC The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a genomic sequence encoding a
XX protein of the invention
XX
XX SQ Sequence 2379 BP; 548 A; 689 C; 539 G; 603 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.08e-179 Length: 2379
Score: 196.00 Matches: 196
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 41.70% Indels: 0
Gaps: 0
DB: 4

US-09-993-966-7 (1-470) x AAL03216 (1-2379)
QY 275 G[ysErProSerVa]a[glInLySerG]luLeuProRoArGThrSerAsnProThArg 294
Db 840 GGCCTCCCTCCGCGGCCGAGAACTGCAAGTCCCTCCCGCCGCACTCCAAATCCACCTCGA 899
QY 295 SerArgSerIegIuProG[ua]a[le]His[IL]ProHisArgLySPROGInLyValAsp 314
Db 900 TCTGCTCCATAGACCGGAAGCATTCACATCCACAGAAAGCCCAAGGGCTGAC 959
QY 315 ProHisSerPheHisPheLeuAspThrProILa[ly]SerGluLeuGInLyArg 334
Db 960 CCGGCTCTCTTCCACTTCTTGACACCCCAATGCCAAGGTCTAGAGCTCCAGCAACGG 1019
QY 335 LeuArgGlyThrGlnaPrgIySerLyHisPheValArgSerProLySaIaGInLyLyS 354
Db 1020 CTCGGGGGACCCGAGGACGAGCAACACTTGTGTGCTCCCCCAAGGCGCCAGGGGCAAG 1079
QY 355 SerValGlyValGlyHisVala[la]ArgGlyValaArgAsnLySPROLeuGlyProAla 374
Db 1080 AGTGTGGGTGTGGCCAGTGGCCAGAGGGGCAAGAAAGCCCTCTGGAGCCGGCC 1139
QY 375 I[le]ProAlaValSerProSerAlaHisLeuAla[le]SerProAlaLeuLeuProSerLeu 394
Db 1140 ATCCCTGCGGTGTCCTCCCTCCGACCACTGGCTGCGAGCCGCGGCTCTCTCCCTA 1199
QY 395 AlaProLeuGlyHisLySlyHisLyHisArgAla[ly]SerGluSerGInLyCyaaArg 414
```


PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.

XX Claim 1; SEQ ID NO 8686; 44pp; English.

XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?docid=20030073623

XX Sequence 427 BP; 84 A; 177 C; 112 G; 53 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.:	3,87e-124	Length:	427
Score:	138.00	Matches:	138
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	29.36%	Indels:	0
DB:	9	Gaps:	0

US-09-993-966-7 (1-470) x ACH21474 (1-427)

Qy 285 LeuProFroArgThrSerAspProThrArgSerArgSerHisGluProGluAlaIleHis 304
Db 4 CTGCCCCCGGACCTCCATCCACATCGATCGCTCCCATGAGCGGAAGCATCATCAC 63
Qy 305 IleProHisArgIysPProGInGlyValAspProAlaSerPheHisPheLeuAspThrPro 324
Db 64 ATCCCAACACGAAAGCCCAAGGCGTGACCGCGCTCTTCCACTTGTGACACCCCA 123
Qy 325 IlAlaIysValSerGluLeuGInGlyArgLeuArgGlyThrGlnAspGlySerIysHis 344
Db 124 ATGCCCAAGGCTCTAGAGCTCCACAAAGGCTCCGGGACCCAGAGGAGCAAGCAC 183
Qy 345 PheValArgSerProIysAlaGInGlyIysSerValGlyValGlyHisValAlaArgGly 364
Db 184 TTGTGAGGTCCGCCAAGGCGCAAGGCAAGTGTGGGTGTGGGCCACGTGGCCAGAGGG 243
Qy 365 AlaArgAsnIysPProPoleuGlyProAlaIleProAlaValSerProSerAlaHisLeu 384
Db 244 GCAAGAAACAGCCCTCTGTGGACCGGCATCCCTCGGTGTGCCCCCTCGCCACCTG 303
Qy 385 AlaAlaSerProAlaLeuLeuProSerLeuAlaProleuGlyHisIysIysValHisHis 404
Db 304 GCTCCAGCCCGGCGCTCTCTCCCTAGGCCCCCTCGGAGCAAGAGACAGACAGAC 363
Qy 405 ArgAlaIysGluSerGInGInGlyIysArgGlyLeuGInAlaProleuAlaSer 422
Db 364 CGAGCCAAAGAGACCAACAGGGCTGCGGGGCTGCAAGACCACTGGCTCA 417

RESULT 7

ID AAS92082 standard; cDNA; 1561 BP.

XX AAS92082;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #27886.

KM Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001MO-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Dmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PsDB; ABG27895.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX Claim 1; SEQ ID NO 27886; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAG64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1561 BP; 396 A; 459 C; 397 G; 309 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.57e-93	Length:	1561
Score:	107.00	Matches:	150
Percent Similarity:	98.68%	Conservative:	0
Best Local Similarity:	98.68%	Mismatches:	1
Query Match:	22.77%	Indels:	2
DB:	5	Gaps:	0

US-09-993-966-7 (1-470) x AAS92082 (1-1561)

Qy 124 LeuGInGlyAspValSerMetGluGluAspSerArgGInGluTrpThrPheThrLeuTyr 143

Db 3 CTCAGTGCACATGTTCATGAGAGACGCGGAGAGATGAGACTTACCTCTAT 62

Qy 144 AspPheAspAsnAsnGlyIysValThrArgGluAspIleThrSerLeuLeuHisThrIle 163

Db 63 GACTTGACAACAACGCAAGGCTCACCCGAGAGACATCACAGCTTGCTGCACACATC 122

Qy 164 TyrGluValVal-AspSerSerValAsnHisSerProThrSerSerIysMetLeuArgVal 183

Db 123 TATGAGGTGC-GGACTCTCTGTCAACCACTCCCAACATCCAGCAATGTGCGGGT 181

QY	183	lyleuethrValAlaProAspGlySerGlnSerIlyValArgSerValIleuValAsnGlnAl	203
Db	182	AAAGTCAACCGTGGCCCCCGAATGGCAGCGACGAAGCAAGAGAGGGCTCTGTCAATCAGGC	241
QY	203	AAAPLeuGlnSerAlaArgProArgAlaGluThrIlyProThrGluAbpLeuArgSerTr	223
Db	242	TGACCTGCAGAGCCGACAGGCCCCCGAGCAAGACCAAGCCCACTGAGGACTCTGGGAGCTG	301
QY	223	PGluIlyAlpGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGluGlnSerG1	243
Db	302	GGAGAAAGACAGCAGCAGCCCGCTCAGGTTCCAGGGGTGACAGCCGCTGGAGCAGTCTGG	361
QY	243	YCyeryrthIshISCyValAspGluAbnGlnIleGluAlaGluArgAsnIstIlyLeuArgLeuAl	263
Db	362	CTGCTACCAACCACTGTGCGTAGATGAGAACTGCAAGAGAGAAACCACTACTTAAATCTCGC	421
QY	263	AGIlyIleGluAenIyTrhSerGlnPheG1yPro	274
Db	422	CGGGATAGAAACTACACGCTCCCAATTGGGCT	455
RESULT 8			
AA578752			
ID	AA578752	standard; cDNA; 1743 BP.	
XX	AA578752;		
AC			
DT	13-FEB-2002	(first entry)	
XX			
DE		DNA encoding novel human diagnostic protein #14556.	
XX			
KW		Human; chromosome mapping; gene mapping; gene therapy; forensic;	
KW		food supplement; medical imaging; diagnostic; genetic disorder; ss.	
OS		Homo sapiens.	
XX			
PN	WO200175067-A2.		
XX			
PD	11-OCT-2001.		
XX			
FP	30-MAR-2001; 2001WO-US008631.		
XX			
PR	31-MAR-2000; 2000US-00540217.		
PR	23-AUG-2000; 2000US-00649167.		
XX			
FA	(HYSE-) HYSEQ INC.		
XX			
P1	Dmanec RT, Liu C, Tang YT;		
XX			
DR	WPI; 2001-639362/73.		
XX			
DR	P-PSDB; ABG14565.		
PT			
XX			
PS	Claim 1; SEQ ID NO 14556; 103bp; English.		
XX			
CC	The invention relates to isolated polynucleotide (I) and polypeptide (II)		
CC	sequences. (I) is useful as hybridisation probes, polymerase chain		
CC	reaction (PCR) primers, oligomers, and for chromosome and gene mapping,		
CC	and in recombinant production of (II). The polynucleotides are also used		
CC	in diagnostics as expressed sequence tags for identifying expressed		
CC	genes. (II) is useful in gene therapy techniques to restore normal		
CC	activity of (I) or to treat disease states involving (II). (II) is		
CC	useful for generating antibodies against it, detecting or quantitating a		
CC	polypeptide in tissue, as molecular weight markers and as a food		
CC	supplement. (II) and its binding partners are useful in medical imaging		
CC	of sites expressing (II). (I) and (II) are useful for treating disorders		
CC	involving aberrant protein expression or biological activity. The		
CC	polypeptide and polynucleotide sequences have applications in		
CC	diagnostics, forensics, gene mapping, identification of mutations		
CC	in		

CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC	coding sequences of the invention. Note: The sequence data for this
CC	patent did not appear in the printed specification, but was obtained in
CC	electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pcr_sequences
XX	
XX	Sequence 1743 BP, 367 A, 562 C, 511 G, 303 T, 0 U, 0 Other,
XX	
XX	Alignment Scores:
XX	Pred. No.: 1,566-59 Length: 1743
XX	Score: 72.00 Matches: 72
XX	Percent Similarity: 100.00% Conservatave: 0
XX	Best Local Similarity: 100.00% Mismatches: 0
XX	Query Match: 15.32% Indels: 0
XX	DB: 5 Gaps: 0
XX	
XX	US-09-993-966-7 (1-470) x AAS78752 (1-1743)
QY	276 SerProSerValAlaGlnLysSerGluLeuProPArgThrSerAsnProThrArgSer 2955
DB	334 TCCCTTCCTGCGCCCAAGTCAGATCGCCCGCAGCTCCATCCATCGATCT 393
QY	296 ArgSerHisGluProGlnAlaIleHisIleProHisArgLysProGlnGlyValAspPro 315
DB	394 CGCTCCCATGAGCCGGAAGCCATCCACATCCACACCGGAAGCCCGAGCGGACCCG 453
QY	316 AlaSerPheHisPheLeuAspThrProIleAlaLysValSerGluLeuGlnArgLysLeu 335
DB	454 GCCTCTTCCACTCTTGTGACACCCCAATGCCCAAGTGTCCAGCTCCAGCAACGAGCTC 513
QY	336 ArgGlyThrGlnAspGlySerLysHisPheValArg 347
DB	514 CGGGCAACCCAGACGGAAGCAAGCACTTTGTGAGG 549
XX	
XX	RESULT 9
XX	AAS92081
XX	ID AAS92081 standard, cDNA, 1743 BP.
XX	AC AAS92081;
XX	DT 13-FEB-2002 (first entry)
XX	DE DNA encoding novel human diagnostic protein #27885.
XX	Human, chromosome mapping; gene mapping; gene therapy; forensic;
XX	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	Homo sapiens.
XX	WO200175067-A2.
XX	PD 11-OCT-2001.
XX	PF 30-MAR-2001, 2001WO-US008631.
XX	PR 31-MAR-2000, 2000US-00540217.
XX	PR 23-AUG-2000, 2000US-00649167.
XX	PA (HYSE-) HYSEQ INC.
XX	Drmanac RT, Liu C, Tang YT,
XX	WPI; 2001-639362/73.
XX	P-PSDB; ABG27894.
XX	New isolated polynucleotide and encoded polypeptides, useful in
XX	diagnostics, forensics, gene mapping, identification of mutations
XX	responsible for genetic disorders or other traits and to assess
XX	biodiversity.
XX	Claim 1; SEQ ID NO 27885, 103pp; English.


```
Db 394 CGCTCCATGAGCCGGAAGCCATCCATCCCAACGGAAGCCCGGCGGACCCG 453
Oy 316 AlaSerPheHisPheLeuAspThrProIleAlaIysValSerGluLeuGlnGlnArgIeu 335
Db 454 GCCTCCTTCACCTTCCTTGAACCCCAATGCCAAGGCTCAGAGCTCCAGCAACGGCTC 513
Oy 336 ArgGlyThrGlnAapGlySerIysHisPheValArg 347
Db 514 CGGGGACCCAGAGCAGGAGCAACACTTGTGTAGG 549

RESULT 11
AAL51522
ID AAL51522 standard; DNA; 148 BP.
XX
AC AAL51522;
XX
DE 24-APR-2003 (first entry)
XX
XX Human Nkd (hmkd) gene exon 8 (coding region).
XX
KM Human; ds; gene therapy; Nkd; hmkd; wnt signalling pathway; cancer;
KM colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer;
KM promoter.
XX
OS Homo sapiens.
XX
PN MO200292832-A2.
XX
PD 21-NOV-2002.
XX
PF 27-NOV-2001; 2001MO-US044092.
XX
PR 27-NOV-2000; 2000US-0252884P.
PR 16-MAY-2001; 2001US-0291109P.
PR 01-OCT-2001; 2001US-0325571P.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Rohan M, Chan V, Yan D;
XX
DR WPI; 2003-129303/12.
XX
XX New human and non-human primate homologues of Nkd protein, and Nkd genes,
XX useful for treating cancer involving aberrant wnt signaling, e.g. colon
XX cancer, head and neck cancer, ovarian cancer, or breast cancer.
XX
PS Example 3; Fig 7; 99pp; English.
XX
XX The invention comprises the amino acid and coding sequence of the human
XX Nkd protein (hmkd) - a regulator of wnt signalling pathways. The hmkd DNA
XX and protein sequences are useful for the treatment of cancer involving
XX aberrant wnt signalling (e.g. colon cancer, head and neck cancer, ovarian
XX cancer and breast cancer). The present DNA sequence represents exon 8
XX (coding region) from the human Nkd gene
XX
SQ Sequence 148 BP; 35 A; 48 C; 38 G; 27 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,06e-38 Length: 148
Score: 49.00 Matches: 49
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.43% Indels: 0
Gaps: 0
DB: 10

US-09-993-966-7 (1-470) x AAL51522 (1-148)

Oy 155 AspIleThrSerLeuLeuHisThrIleTyrgIuValAlaPssSerSerValAsnHisSer 174
Db 1 GACATCACGAGCTTGACACACCATCTATGAGGTGGAGCTCCTGTCTCAACCACTCC 60
Oy 175 ProThrSerSerIysMetLeuArgValIleLeuThrValAlaProAapGlySerGlnSer 194
```

```
Db 61 CCAACATCCAGCAAGATGCTGGGGTAAAGCTCACCGTGGCCCCGATGGCAGCCAGAC 120
Oy 195 LysArgSerValIeuValaIngIala 203
Db 121 AAGAGAGCGCTCTTGTCAATCAGGCT 147

RESULT 12
AAL51512
ID AAL51512 standard; DNA; 1401 BP.
XX
AC AAL51512;
XX
DE 24-APR-2003 (first entry)
XX
XX Mouse Nkd protein gene sequence.
XX
KM Mouse; gene; ds; gene therapy; Nkd; hmkd; wnt signalling pathway; cancer;
KM colon cancer; head cancer; neck cancer; ovarian cancer; breast cancer;
KM murine.
XX
OS Mus sp.
XX
PN MO200292832-A2.
XX
PD 21-NOV-2002.
XX
PF 27-NOV-2001; 2001MO-US044092.
XX
PR 27-NOV-2000; 2000US-0252884P.
PR 16-MAY-2001; 2001US-0291109P.
PR 01-OCT-2001; 2001US-0325571P.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Rohan M, Chan V, Yan D;
XX
DR WPI; 2003-129303/12.
XX
XX New human and non-human primate homologues of Nkd protein, and Nkd genes,
XX useful for treating cancer involving aberrant wnt signaling, e.g. colon
XX cancer, head and neck cancer, ovarian cancer, or breast cancer.
XX
PS Example 2; Fig 4; 99pp; English.
XX
XX The invention comprises the amino acid and coding sequence of the human
XX Nkd protein (hmkd) - a regulator of wnt signalling pathways. The hmkd DNA
XX and protein sequences are useful for the treatment of cancer involving
XX aberrant wnt signalling (e.g. colon cancer, head and neck cancer, ovarian
XX cancer and breast cancer). The present DNA sequence represents the mouse
XX Nkd gene
XX
SQ Sequence 1401 BP; 336 A; 430 C; 416 G; 219 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2.6e-37 Length: 1401
Score: 49.00 Matches: 49
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.43% Indels: 0
Gaps: 0
DB: 10

US-09-993-966-7 (1-470) x AAL51512 (1-1401)

Oy 131 GluGluAspSerArgGlnGluTrrPheThrLeuTyraPphAspAsnAngIyls 150
Db 391 GAGGAGGACAGCCGGGAGAGTGACTTTCACCTCTATATACCTTCGACACATGGGAAA 450
Oy 151 ValThrArgGluAspIleThrSerLeuLeuHisThrIleTyrgIuValAlaPssSerSer 170
Db 451 GTGACCCGTAAGGACATTAACACTTGCTGCATACATCATATGAAGGTGTGACTCTCT 510
Oy 171 ValAsnHisSerProThrSerSerIys 179
```

Db 511 GTGAACCATTCCTCCCACTACAGCAAG 537

RESULT 13
AAH28343
ID AAH28343 standard; DNA; 1416 BP.
XX
AC AAH28343;
XX
DT 05-SEP-2001 (first entry)
XX
DE Nucleotide sequence of a murine mnkd protein.
XX
KM Dishevelled-associated protein 1A; DAP1A; mnkd; dishevelled protein;
KM Wnt signalling; colon cancer; cancer; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 1. 1416
FT /tag= a
FT /product= "mnkd"
XX
XX MO200144279-A2.
XX
PD 21-JUN-2001.
XX
PF 05-DEC-2000; 2000MO-US032986.
XX
PR 17-DEC-1999; 99US-0172434P.
XX
XX (CHIR) CHIRON CORP.
XX
PI Yan D, Williams LT;
XX
XX WPI; 2001-425440/45.
DR P-PSDB; AAB84645.
XX
XX Novel mammalian dishevelled-associated proteins, mnkd and DAP1A, useful
PT for inhibiting Wnt signaling in mammalian cells, and thus for treating
PT colon cancer.
XX
XX Claim 2; Fig 1; 67pp; English.
XX
XX The present sequence encodes a murine mnkd protein. The specification
CC describes mnkd and dishevelled-associated protein 1A (DAP1A) proteins,
CC that interact with mammalian dishevelled protein. mnkd is useful for
CC inhibiting Wnt signaling in mammalian cells, and thus for treating colon
CC cancer. mnkd is also useful for activating the JNK pathway. mnkd and
CC DAP1A are also useful for screening drugs that are useful for treating
CC cancer
XX
XX SQ Sequence 1416 BP; 339 A; 434 C; 419 G; 224 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2.62e-37 Length: 1416
Score: 49.00 Matches: 49
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.43% Indels: 0
Gaps: 0
DB: 4

US-09-993-966-7 (1-470) x AAH28343 (1-1416)

QY 131 GUGUuapSerAraGingIuTPThrPheThrLeuTyraSpheaspasnaNglyLys 150
Db 391 GAGGAGGACAGCCGCGCAAGAGTGACTTCACTCATATGACTTCGACCAACATGGCAAA 450
QY 151 ValThrArgGluuapIleThrSerLeuLeuHisThrIleTyrgIuValValAspSerSer 170
Db 451 GTGACCCGTGAGGACATTACAGCTTGCTGCATACCACTATGAAGTGGTGAAGTCTCTCT 510
QY 171 ValAsnHisSerProThrSerSerLys 179

Db 511 GTGAACCATTCCTCCCACTACAGCAAG 537

RESULT 14
ABQ73820
ID ABQ73820 standard; cDNA; 1950 BP.
XX
AC ABQ73820;
XX
DT 07-OCT-2002 (first entry)
XX
DE Human colon specific nucleic acid (CSNA) SEQ ID NO:126.
XX
KM Human; colon specific nucleic acid; colon specific polypeptide; CSP;
KM CSNA; colon specific gene; CSG; colon cancer; gene therapy; vaccine;
KM cytoabatic; gene; ss.
XX
OS Homo sapiens.
XX
PN MO200248370-A2.
XX
PD 20-JUN-2002.
XX
PF 30-OCT-2001; 2001MO-US051341.
XX
PR 31-OCT-2000; 2000US-0244717P.
XX
XX (DIAD-) DIADEXUS INC.
XX
PI Sun Y, Recipon H, Ghosh MG, Liu C;
XX
XX WPI; 2002-583520/62.
XX
XX Colon specific polypeptides and polynucleotides useful for detecting,
PT diagnosing, monitoring, treating, staging and predicting cancers in
PT humans having cancer and non-cancerous colon disease.
XX
XX Claim 1; Page 199; 243pp; English.
XX
XX ABQ73695 to ABQ73841 represent human colon specific nucleic acid (CSNA)
CC sequences, and ABP51826 to ABP51928 represent human colon specific
CC polypeptide (CSP) sequences from the present invention. CSNA and CSP
CC sequences have cytostatic activity, and can be used in gene therapy,
CC antisense therapy and in vaccines. CSNA and CSP sequences can be used for
CC diagnosing and monitoring the presence and metastases of colon cancer in
CC a patient, by determining an amount of CSP or CSNA in a sample of a
CC patient, and comparing it to the amount of colon specific marker in a
CC normal control, where a difference in the amount of the nucleic acid or
CC the polypeptide in the sample compared to that of normal control is
CC associated with presence of colon cancer. CSP and CSNA sequences can be
CC used for producing engineered colon tissue for treatment and research.
CC CSNA sequences are useful for producing transgenic animals and cells and
CC also in gene therapy
XX
XX SQ Sequence 1950 BP; 459 A; 494 C; 530 G; 467 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2.89e-34 Length: 1950
Score: 46.00 Matches: 46
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.79% Indels: 0
Gaps: 0
DB: 6

US-09-993-966-7 (1-470) x ABQ73820 (1-1950)

QY 20 GUYAspSerPheAlaValISerAlaATPAlaArgySgIyIleGluGluTPPileGly 39
Db 838 GGTGACAGCTTCCCGCGAGCCTGCTGGCTCGGAAGGCAATCGAGATGATGGG 897
QY 40 ArgGlnArgCysPArgIyGlyValISerGlyProArgGlnLeuArgLeuAlaGlyThrIle 59
Db 898 AGACAGCGCTGCCCGGCGGTGTCTCGGAGACCCGACAGCTGCGGTGGCGGCGACCATTA 957

QY 60 GlyArgSerThrArgGlu 65
 DB 958 GGCGAAGACACCGGAA 975
 RESULT 15
 ID ACH74795 standard; DNA, 598 BP.
 AC ACH74795;
 XX 29-JUL-2004 (first entry)
 DT 29-JUL-2004 (first entry)
 XX Human genome derived single exon probe #7990.
 DE Human; probe; ss; gene expression; single exon probe; microarray;
 KW alternative splicing event; genomic alteration.
 XX Homo sapiens.
 OS US2003194704-A1.
 XX 16-OCT-2003.
 PD 03-APR-2002; 2002US-00029386.
 PF 03-APR-2002; 2002US-00029386.
 PR 03-APR-2002; 2002US-00029386.
 XX (PENN/) PENN S G.
 PA (RANK/) RANK D R.
 PA (HANZ/) HANZEL D K.
 PI Penn SG, Rank DR, Hanzel DK;
 DR MPI; 2004-119264/12.
 XX New human genome-derived single exon nucleic acid probes useful for human
 PT gene expression analysis, for identifying or characterizing alternative
 PT splicing events, for assessing genomic alterations or as tools for
 PT surveying tissues.
 XX Claim 15; SEQ ID NO 7990; 80bp; English.
 CC The invention relates to a nucleic acid probe for measuring human gene
 CC expression, comprising any of the 27,400 fully defined nucleotide
 CC sequences in the specification, or their complements or fragments, and
 CC encoding at least 8 amino acids of any of the 688 amino acid sequences
 CC fully defined in the specification. The probe is a single exon probe that
 CC hybridizes under high stringency conditions to a nucleic acid molecule
 CC expressed in human cells or tissues. Also included are a spatially-
 CC addressable set of single exon nucleic acid probes for measuring human
 CC gene expression (comprising a plurality of single exon nucleic acid
 CC probes cited above, where each of the plurality of probes is separately
 CC and addressably isolatable or amplifiable from the plurality), a single
 CC exon microarray for measuring human gene expression, a method of
 CC measuring human gene expression, a vector comprising the single exon
 CC probe cited above, an ORF-encoded peptide comprising at least 8
 CC contiguous amino acids of any of the above-mentioned amino acid
 CC sequences (optionally with conservative amino acid substitutions), an
 CC isolated antibody that binds specifically to a peptide cited above,
 CC methods of selling and/or licensing single exon probes or microarrays to
 CC a customer desiring to measure gene expression, a method of providing
 CC human gene expression data by subscription, and a computer-readable
 CC storage medium which contains a database having a plurality of records
 CC (each record including data on the expression of a single exon probe
 CC cited above). The probe, methods and apparatus are useful in gene
 CC expression analysis. The probes may be used as tools for surveying
 CC tissues to detect the presence of expressed messages that contain their
 CC specific exon, or in constructing genome-derived single exon microarrays.
 CC In addition, the probes are used in identifying and characterizing
 CC alternative splicing events, in detecting and characterizing gross
 CC alterations in the genomic locus that includes their exon, in assessing
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,

CC or in expressing the ORF-encoded peptide. The present sequence is a human
 CC single exon probe of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?docid=20030194704
 XX
 SQ Sequence 598 BP; 84 A; 232 C; 162 G; 120 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 8,776-34 Length: 598
 Score: 45.00 Matches: 45
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 9.57% Indels: 0
 DB: 12 Gaps: 0
 US-09-993-966-7 (1-470) x ACH74795 (1-598)
 QY 20 GlyArgSerPheAlaValSerAlaIleTPAlaArgGlyGlyIleGluGluTyrPileGly 39
 DB 229 GGTGACAGCTTCCGCGGAGCGGCTGCGGCTCGGAGGCGCATCGAGGTGATCGGG 288
 QY 40 ArgGlnArgCysProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIle 59
 DB 289 AGACAGGCGTCCCGGCGGCTGTCTCGGACCCCGACAGCTGCGGCGGCGGACCATTA 348
 QY 60 GlyArgSerThrArg 64
 DB 349 GGCGAAGACACCGG 363

Search completed: December 30, 2004, 09:47:45
 Job time: 860 secs

THIS PAGE IS BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 30, 2004, 09:00:11 ; Search time 142 Seconds
(without alignments)
2352.611 Million cell updates/sec

Title: US-09-993-966-7

Perfect score: 470

Sequence: 1 MGKLSKPAVCKKRESPEG.....RHHHHHHHHHHHHFYQT 470

Scoring table: (OLIGO)
Xgapop 60.0, Xgapext 60.0
Xgapop 60.0, Xgapext 60.0
Xgapop 6.0, Xgapext 7.0
Delop 6.0, Delext 7.0

Searched: 824507 seqs, 355394441 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1643622

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xip
-O/cgn2_1/US993966/runat_29122004_130830_9412/app_query.fasta_1.647
-Database Patents NA -OPMT=fastap -SUFFIX=olig.rnt -MINMATCH=0.1 -IOPTCL=0
-LOOPTXT=0 -UNITS=BITS -START=1 -END=1 -MATRIX=oligo -TRANS=humand0.cdi
-LIST=45 -DOCLIN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0993966 @CEN 1.128 @runat_29122004_130830_9412 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FCAPOP=6
-FCAPEXT=7 -XGAPOP=60 -XGAPEXT=60 -DELOP=6 -DELEXT=7

Database: Issued Patents NA.*

- 1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PTCUS.COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	383	81.5	1438	4	US-09-506-066E-5
2	45	9.6	1731	4	US-09-506-066E-3
3	17	3.6	1285	4	US-09-506-066E-7
4	17	3.6	1307	4	US-09-506-066E-9
5	8	1.7	75	2	US-08-447-169A-125
6	8	1.7	75	2	US-08-233-012C-125
7	8	1.7	75	4	US-09-860-474-125
8	8	1.7	428	4	US-09-641-638-497
9	8	1.7	428	4	US-10-170-097-497
10	8	1.7	449	4	US-09-976-594-405
11	8	1.7	455	4	US-09-621-976-1583
12	8	1.7	458	4	US-09-270-767-3943

13	8	1.7	458	4	US-09-270-767-19325	Sequence 19225, A
14	8	1.7	482	4	US-09-270-767-5688	Sequence 5698, Ap
15	8	1.7	482	4	US-09-270-767-20980	Sequence 20980, A
16	8	1.7	501	4	US-09-583-110-1646	Sequence 1646, Ap
17	8	1.7	531	4	US-09-621-976-586	Sequence 586, App
18	8	1.7	533	4	US-09-270-767-27303	Sequence 27303, A
19	8	1.7	639	4	US-09-270-767-14169	Sequence 14169, A
20	8	1.7	683	4	US-09-270-767-11684	Sequence 11684, A
21	8	1.7	732	4	US-09-248-796A-1692	Sequence 1692, Ap
22	8	1.7	924	4	US-09-270-767-87	Sequence 87, Appl
23	8	1.7	924	4	US-09-270-767-15369	Sequence 15369, A
24	8	1.7	1155	3	US-09-439-313-373	Sequence 373, App
25	8	1.7	1155	3	US-09-352-616A-373	Sequence 373, App
26	8	1.7	1155	4	US-09-289-198-301	Sequence 301, App
27	8	1.7	1155	4	US-09-636-215-373	Sequence 373, App
28	8	1.7	1155	4	US-09-685-166A-373	Sequence 373, App
29	8	1.7	1155	4	US-09-429-755-301	Sequence 301, App
30	8	1.7	1155	4	US-09-679-426-373	Sequence 373, App
31	8	1.7	1225	4	US-09-270-767-10546	Sequence 10546, A
32	8	1.7	1366	4	US-09-149-476A-49	Sequence 49, Appl
33	8	1.7	1512	3	US-09-439-313-368	Sequence 368, App
34	8	1.7	1512	3	US-09-662-451-294	Sequence 294, App
35	8	1.7	1512	3	US-09-352-616A-368	Sequence 368, App
36	8	1.7	1512	4	US-09-636-215-368	Sequence 368, App
37	8	1.7	1512	4	US-09-685-166A-368	Sequence 368, App
38	8	1.7	1512	4	US-09-429-755-294	Sequence 294, App
39	8	1.7	1512	4	US-09-679-426-368	Sequence 368, App
40	8	1.7	1754	4	US-09-248-796A-4659	Sequence 4659, Ap
41	8	1.7	1787	2	US-08-808-982-2	Sequence 2, Appl
42	8	1.7	1787	3	US-09-306-902A-2	Sequence 2, Appl
43	8	1.7	1853	3	US-09-439-313-369	Sequence 369, App
44	8	1.7	1853	3	US-09-062-451-295	Sequence 295, App
45	8	1.7	1853	3		

ALIGNMENTS

RESULT 1
US-09-506-066E-5
Sequence 5, Application US/09506066E
Patent No. 6630323
GENERAL INFORMATION:
APPLICANT: Scott, Matthew
APPLICANT: Wharton, Keith
TITLE OF INVENTION: Naked Cuticle Genes and their Uses
FILE REFERENCE: STAN-121
CURRENT APPLICATION NUMBER: US/09/506, 066E
CURRENT FILING DATE: 2000-02-17
PRIOR APPLICATION NUMBER: 60/120, 646
PRIOR FILING DATE: 1999-02-17
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 1438
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (8) ..(1418)
OTHER INFORMATION: Nkd1 coding sequence
US-09-506-066E-5

Alignment Scores:

Pred. No.: 0
Score: 383.00
Percent Similarity: 99.58%
Best Local Similarity: 99.58%
Query Match: 81.49%
DB: 4
US-09-993-966-7 (1-470) x US-09-506-066E-5 (1-1438)

Length: 1438
Matches: 469
Conservative: 1
Mismatch: 1
Indels: 2
Gaps: 0

```

Qy 1 MetGlyblyseuHiserlybProAlaValaCylsArgrArgluserProglugly 20
Db 8 ATGGGGAATCTTCACTCCAAACCCGCGCTGTGCAACCCAGAGAGAGCCCGAAGGT 67
Qy 21 AspSerPheAlaValSerAlaAlaTPAlaArglygllyleglugluprilegylArg 40
Db 68 GACAGCTTCGCGCTGAGCGCTGCTGGGCTCGAAGGCGATCGAGTGGATCGGGAGA 127
Qy 41 GlnArgCybProgllygllyValSerGlyProArglInleuAgluAgllyThllegly 60
Db 128 CAGGCTGCGCGCGGCTGTCTCGGAGCCCGACAGCTGCGGTGGCGGCACCATAGGC 187
Qy 61 ArgSerThrArgGluLeuValGlyAspValLeuArgbPThrLeuSerGluGluGlu 80
Db 188 CGAAGCACCCGGAGCTGTGGCGACCTGTGAGACACGCTCAGCGAGAAAGAGAG 247
Qy 81 AspArgPheArgLeuGluValAlaLeuProProgllybPThrArgGlyLeuGly 100
Db 248 GACGACTTTCGGCTGGAAGTGGCTGCTCTGAGAAAGACTGACGGGCTGGCAGCGGA 307
Qy 101 AspGluValblybMetGluArgValSerGluProCybProgllySerlyblybGluLeu 120
Db 308 GATAGAGAAGAGATGAGAGAGTGAAGCACTGCGCCAGGCTCCAAAGACAGCTTAG 367
Qy 121 PheGluGluLeuGlnCybAspValSerMetGluGluAspSerArgGlnGluTrpThrPhe 140
Db 368 TTTCGAAGCTCCAGTGCAGCTGTCTCATGAGAGAGACAGCCGCGAGACTGACCTTC 427
Qy 141 ThrLeuTyArgPheAspAsnAsnGlylyValThrArgLysApplleThrSerLeuLeu 160
Db 428 ACCGTATGACTTTGACCAACACGCGAAGGTCAACCGAGAGACATACACAGCTTCTG 487
Qy 161 HisThrLeuTyArgValValaAspSerSerValAsnHisSerProThrSerSerlyMet 180
Db 488 CACGCCATCATATGAGGTGTGAGTCTCTGTCTCAACCACTCCCAACATCCAGCAATG 547
Qy 181 LeuArgValblybLeuThrValAlaProAspGlySerGlnSerlybAspSerValLeuVal 200
Db 548 CTGGGGGTAAAGCTCACCGTGCGCCCGATGGCGAGCCAGACAGAGAGAGGCTCTTGT 607
Qy 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgLagluThrybProThrgluAspLeu 220
Db 608 AATCAGGCTGACCTGCAGAGCGCAAGGCCCGAGCAGAGACCAAGCCACATGAGACTG 667
Qy 221 ArgSerTrpGluValblybGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
Db 668 CGGAGCTGGAGAAAGAGAGCGAGCGCCGCTCAGGTTCAGGGGTGACGCGCTGAG 727
Qy 241 GlnSerGlyCybTyRHisbCybValaAspGluAsnilegluArgAsnHisTyRLeu 260
Db 728 CAGTCTGGCTCTACCAACATGCTGATGAGAAACATCGAGAGAGAAACCACTACTTA 787
Qy 261 AspLeuAlaglylegluAsnTyRThrSerGlnPheGlyProgllySerProSerValAla 280
Db 788 GATCTCCCGGAGTAGAAACTACACGCTCCCAATTTGGGCTGCTCCCTTCCTGGGCC 847
Qy 281 GlnlybSerGluLeuProProArgThrSerAspProThrArgSerArgbPThrSer 300
Db 848 CAGAGTCAAGACTGCGCCCGCACCTTCAATCCCACTGACTGCTCCCATGAGGCC 906
Qy 300 OGAlaAlaIleHisleProHisArglybProGlnGlyValaAspProLasePheHisPhe 320
Db 907 GGAAGCATTCACATCCACACCGAAAGCCCAAGGGGTGAGCCCGGCTCTTCCACTT 966
Qy 320 eLeuAspThrProIleAlaValSerGluLeuGlnGlnArgLeuArglyThrglnAs 340
Db 967 CTTTGAACACCCCAATCGCAAGTCTCAGAGCTCCAGCAAGGCTCGGGGCACTCAGGA 1026
Qy 340 PGIySerlybHisPheValaArgSerProValbGlnGlylybSerValGlyValGly 360
Db 1027 CGGGAGCAAGCACTTTGTGAGGTCCCGCAAGGCCAGAGGCAAGGTGTGGTGGGCCA 1086
Qy 360 sValAlaArgGlyValaArgblybProProLeuGlyProAlaIleProAlaValSerPr 380

```

```

Db 1087 CTGGCCAGAGGGGGCAAGAAACAAGCCCTCTGGAGCCGCGCATCTCGGTGTCC 1146
Qy 380 oSerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlybly 400
Db 1147 CTCGGCCACCTGCTGCTGCGAGCCCGGCTCTCTCCCTCTGAGCCCTCCGGCAAA 1206
Qy 400 slybHisblybArgAlaValbGluSerGlnGlnGlyCybArgGlyLeuGlnAlaProle 420
Db 1207 GAAGCACAGACCGAGCCCAAGAGAGCCAGAGGCTGCGGGGCTGCGAGCCACT 1266
Qy 420 uAlaSerGlyGlyProValLeuGlyValArgGluHisLeuArgGluLeuProAlaLeuVala 440
Db 1267 GGCCTGAGTGGCGCTGTCTGTGGGGGGAGACCTGCGGAGCTGCGGCTTGGTGT 1326
Qy 440 lTyRGluserGlnAlaglyGlnProValGlnArgHisGluHisbHisblyb 460
Db 1327 GTATGAGAGCCAGGCGGGGAGCGGCTCCAGAGACATGAGCACACACCATGAA 1386
Qy 460 sHisbHisbTyRHisbPheTyRGlInhr 470
Db 1387 TCACCACTTACCACTTCTTACAGACA 1417

RESULT 2
US-09-506-066E-3
; Sequence 3 Application US/09506066E
; Patent No. 6630323
; GENERAL INFORMATION:
; APPLICANT: Scott, Matthew
; APPLICANT: Wharton, Keith
; APPLICANT: Zeng, Wenlin
; TITLE OR INVENTION: Naked Cuticle Genes and their Uses
; FILE REFERENCE: STRA-121
; CURRENT APPLICATION NUMBER: US/09/506, 066E
; CURRENT FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/120, 646
; PRIOR FILING DATE: 1999-02-17
; NUMBER OF SEQ. ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (140)...(1553)
; OTHER INFORMATION: Nkdl coding sequence
US-09-506-066E-3

Alignment Scores:
Pred. No.: 2,28e-33 Length: 1731
Score: 45.00 Matches: 45
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.57% Indels: 0
DB: 4 Gaps: 0

US-09-993-966-7 (1-470) x US-09-506-066E-3 (1-1731)
Qy 1 MetGlyblyseuHiserlybProAlaValaCylsArgrArgluserProglugly 20
Db 140 ATGGGGAATCTTCACTCCAAACCCGCGCTGTGCAACCCAGAGAGAGCCCGAAGGT 199
Qy 21 AspSerPheAlaValSerAlaAlaTPAlaArglygllyleglugluprilegylArg 40
Db 200 GACAGCTTCGCGCTGAGCGCTGCTGGGCTCGAAGGCGATCGAGTGGATCGGGAGG 259
Qy 41 GlnArgCybProglly 45
Db 260 CAGCGCTGTCCAGGC 274

RESULT 3
US-09-506-066E-7

```

Sequence 7, Application US/09506066E
 Patent No. 6630323
 GENERAL INFORMATION:
 APPLICANT: Scott, Matthew
 APPLICANT: Wharton, Keith
 APPLICANT: Zeng, Wenlin
 TITLE OF INVENTION: Naked Cuticle Genes and their Uses
 FILE REFERENCE: STAN-121
 CURRENT APPLICATION NUMBER: US/09/506,066E
 CURRENT FILING DATE: 2000-02-17
 PRIOR APPLICATION NUMBER: 60/120,646
 PRIOR FILING DATE: 1999-02-17
 NUMBER OF SEQ ID NOS: 15
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 7
 LENGTH: 1285
 TYPE: DNA
 ORGANISM: Mus musculus
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (2)...(1282)
 OTHER INFORMATION: Nkd2 coding sequence
 US-09-506-066E-7

Alignment Scores:
 Pred. No.: 9.31e-07 Length: 1285
 Score: 17.00 Matches: 17
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.62% Indels: 0
 DB: 4 Gaps: 0

US-09-993-966-7 (1-470) x US-09-506-066E-7 (1-1285)

Qy 254 GIUARGAASHISTYRLAASPLEUA1AG1Y11EG1UASNTYRTHSR 270
 Db 644 GAGCGGAAACCACTACCTGACCTTGCGCATCGAGAACTACACATCT 694

RESULT 4
 US-09-506-066E-9
 Sequence 9, Application US/09506066E
 Patent No. 6630323
 GENERAL INFORMATION:
 APPLICANT: Scott, Matthew
 APPLICANT: Wharton, Keith
 APPLICANT: Zeng, Wenlin
 TITLE OF INVENTION: Naked Cuticle Genes and their Uses
 FILE REFERENCE: STAN-121
 CURRENT APPLICATION NUMBER: US/09/506,066E
 CURRENT FILING DATE: 2000-02-17
 PRIOR APPLICATION NUMBER: 60/120,646
 PRIOR FILING DATE: 1999-02-17
 NUMBER OF SEQ ID NOS: 15
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 9
 LENGTH: 1307
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-506-066E-9

Alignment Scores:
 Pred. No.: 9.46e-07 Length: 1307
 Score: 17.00 Matches: 17
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.62% Indels: 0
 DB: 4 Gaps: 0

US-09-993-966-7 (1-470) x US-09-506-066E-9 (1-1307)

Qy 254 GIUARGAASHISTYRLAASPLEUA1AG1Y11EG1UASNTYRTHSR 270
 Db 364 GAGCGGAAACCACTACCTGACCTTGCGCATCGAGAACTACACATCT 414

RESULT 5
 US-08-447-169A-125/c
 Sequence 125, Application US/08447169A
 Patent No. 5811533
 GENERAL INFORMATION:
 APPLICANT: JANIC, N. and GOLD, L.
 TITLE OF INVENTION: HIGH-AFFINITY OLIGONUCLEOTIDE
 TITLE OF INVENTION: LIGANDS TO VASCULAR ENDOTHELIAL
 NUMBER OF INVENTION: GROWTH FACTOR (VEGF)
 NUMBER OF SEQUENCES: 242
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Swanson & Bratschun, L.L.C.
 STREET: 8400 E. Prentice Place, Suite 200
 CITY: Englewood
 STATE: Colorado
 COUNTRY: USA
 ZIP: 80111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MG storage
 COMPUTER: IBM compatible
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: WordPerfect 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/447,169A
 FILING DATE: 19-MAY-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/233,012
 FILING DATE: 25-APRIL-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/205,515
 FILING DATE: 03-MARCH-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/964,624
 FILING DATE: 21-OCTOBER-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/714,131
 FILING DATE: 10-JUNE-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/536,428
 FILING DATE: 11-JUNE-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Barry J. Swanson
 REGISTRATION NUMBER: 33,215
 REFERENCE/DOCKET NUMBER: NEX14
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (303) 793-3333
 TELEFAX: (303) 793-3433
 INFORMATION FOR SEQ ID NO: 125:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 75 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 FEATURE:
 OTHER INFORMATION: All C's are 2'-NH2 cytosine
 FEATURE:
 OTHER INFORMATION: All U's are 2'-NH2 uracil
 US-08-447-169A-125

Alignment Scores:
 Pred. No.: 24.6 Length: 75
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.70% Indels: 0
 DB: 1 Gaps: 0

US-09-993-966-7 (1-470) x US-08-447-169A-125 (1-75)

Qy 387 SerProlaLeuueProSerLeu 394
 ||||||||||||||||||||||||||||

Db 43 TCACCTGCCCTCTTCCTCCTTG 20

RESULT 6
US-08-233-012C-125/c

; Sequence 125, Application US/08233012C
; Patent No. 5849479

GENERAL INFORMATION:

APPLICANT: JANJIC, N. and GOLD, L.

TITLE OF INVENTION: HIGH-AFFINITY OLIGONUCLEOTIDE

TITLE OF INVENTION: LIGANDS TO VASCULAR

TITLE OF INVENTION: ENDOTHELIAL GROWTH FACTOR

NUMBER OF SEQUENCES: 146

CORRESPONDENCE ADDRESS:

ADDRESSEE: Swanson & Bratschun, L.L.C.

STREET: 8400 E. Prentice Place, Suite 200

CITY: Englewood

STATE: Colorado

COUNTRY: USA

ZIP: 80111

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MG

COMPUTER: IBM compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/233,012C

FILING DATE: 25-APRIL-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/714,131

FILING DATE: 10-JUNE-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/536,428

FILING DATE: 11-JUNE-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/964,624

FILING DATE: 21-OCTOBER-1992

ATTORNEY/AGENT INFORMATION:

NAME: Barry J. Swanson

REGISTRATION NUMBER: 33,215

REFERENCE/DOCKET NUMBER: NEX14

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 793-3333

TELEFAX: (303) 793-3433

INFORMATION FOR SEQ ID NO: 125:

SEQUENCE CHARACTERISTICS:

LENGTH: 75 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

FEATURE:

OTHER INFORMATION: All C's are 2'-NH2 cytosine

FEATURE:

OTHER INFORMATION: All U's are 2'-NH2 uracil

US-08-233-012C-125

Alignment Scores:

Pred. No.: 24.6

Score: 8.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 1.70%

DB: 2

US-09-993-966-7 (1-470) x US-08-233-012C-125 (1-75)

Qy 387 SerProAlaLeuLeuProSerLeu 394

Db 43 TCACCTGCCCTCTTCCTCCTTG 20

RESULT 7

US-09-860-474-125/c

; Sequence 125, Application US/09860474

; Patent No. 666252

GENERAL INFORMATION:

APPLICANT: GOLD, L. and JANJIC, N.

TITLE OF INVENTION: HIGH-AFFINITY OLIGONUCLEOTIDE LIGANDS

TITLE OF INVENTION: TO VASCULAR ENDOTHELIAL GROWTH

NUMBER OF SEQUENCES: 242

CORRESPONDENCE ADDRESS:

ADDRESSEE: Swanson & Bratschun, L.L.C.

STREET: 1745 Shea Center Drive, Suite 330

CITY: Highlands Ranch

STATE: Colorado

COUNTRY: USA

ZIP: 80128

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MG storage

COMPUTER: IBM compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/860,474

FILING DATE: 18-May-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/156,824

FILING DATE: 18-SEPTEMBER-1998

APPLICATION NUMBER: 08/447,169

FILING DATE: 19-MAY-1995

APPLICATION NUMBER: 08/233,012

FILING DATE: 25-APRIL-1994

APPLICATION NUMBER: 08/205,515

FILING DATE: 03-MARCH-1994

APPLICATION NUMBER: 07/964,624

FILING DATE: 21-OCTOBER-1992

APPLICATION NUMBER: 07/714,131

FILING DATE: 10-JUNE-1991

APPLICATION NUMBER: 07/536,428

FILING DATE: 11-JUNE-1990

ATTORNEY/AGENT INFORMATION:

NAME: Barry J. Swanson

REGISTRATION NUMBER: 33,215

REFERENCE/DOCKET NUMBER: NEX14/CIP-CON

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 793-3333

TELEFAX: (303) 793-3433

INFORMATION FOR SEQ ID NO: 125:

SEQUENCE CHARACTERISTICS:

LENGTH: 75 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

FEATURE:

OTHER INFORMATION: All C's are 2'-NH2 cytosine

FEATURE:

OTHER INFORMATION: All U's are 2'-NH2 uracil

SEQUENCE DESCRIPTION: SEQ ID NO: 125:

US-09-860-474-125

Alignment Scores:

Pred. No.: 24.6

Score: 8.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 1.70%

DB: 4

US-09-993-966-7 (1-470) x US-09-860-474-125 (1-75)

Qy 387 SerProAlaLeuLeuProSerLeu 394

Db 43 TCACCTGCCCTCTTCCTCCTTG 20


```
RESULT 8
US-09-641-638-497
; Sequence 497, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; FILE REFERENCE: GENSET 051CPI
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 497
; LENGTH: 428
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 281
; OTHER INFORMATION: 10-15-281 : polymorphic base G or T
; NAME/KEY: misc_binding
; LOCATION: 261..280
; OTHER INFORMATION: 10-15-281.misl, potential
; NAME/KEY: misc_binding
; LOCATION: 282..301
; OTHER INFORMATION: 10-15-281.mis2, potential complement
; NAME/KEY: primer_bind
; LOCATION: 1..20
; OTHER INFORMATION: upstream amplification primer
; NAME/KEY: primer_bind
; LOCATION: 409..428
; OTHER INFORMATION: downstream amplification primer, complement
; NAME/KEY: misc_binding
; LOCATION: 269..293
; OTHER INFORMATION: 10-15-281 potential probe
US-09-641-638-497

Alignment Scores:
Pred. No.: 128 Length: 428
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.70% Indels: 0
DB: 4 Gaps: 0

US-09-993-966-7 (1-470) x US-09-641-638-497 (1-428)
QY 388 ProAlaLeuProSerLeuAla 395
Db 89 CCAGCCTGCTCCCTTCTCTGACC 112

RESULT 9
US-10-170-097-497
; Sequence 497, Application US/10170097
; Patent No. 6794143
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
```

```
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; FILE REFERENCE: GEN-T114XC2D1
; CURRENT APPLICATION NUMBER: US/10/170,097
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: US 09/641,638
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 497
; LENGTH: 428
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 281
; OTHER INFORMATION: 10-15-281 : polymorphic base G or T
; NAME/KEY: misc_binding
; LOCATION: 261..280
; OTHER INFORMATION: 10-15-281.misl, potential
; NAME/KEY: misc_binding
; LOCATION: 282..301
; OTHER INFORMATION: 10-15-281.mis2, potential complement
; NAME/KEY: primer_bind
; LOCATION: 1..20
; OTHER INFORMATION: upstream amplification primer
; NAME/KEY: primer_bind
; LOCATION: 409..428
; OTHER INFORMATION: downstream amplification primer, complement
; NAME/KEY: misc_binding
; LOCATION: 269..293
; OTHER INFORMATION: 10-15-281 potential probe
US-10-170-097-497

Alignment Scores:
Pred. No.: 128 Length: 428
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.70% Indels: 0
DB: 4 Gaps: 0

US-09-993-966-7 (1-470) x US-10-170-097-497 (1-428)
QY 388 ProAlaLeuProSerLeuAla 395
Db 89 CCAGCCTGCTCCCTTCTCTGACC 112

RESULT 10
US-09-976-594-405
; Sequence 405, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
```

```
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 405
; LENGTH: 449
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 199286.3
; LOCATION: 397
; NAME/KEY: unanure
; OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-405

Alignment Scores:
Pred. No.: 134          Length: 449
Score: 8.00           Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.70%      Indels: 0
DB: 4                  Gaps: 0

US-09-993-966-7 (1-470) x US-09-976-594-405 (1-449)

QY 384 LeuAlaAlaSerProAlaLeuLeu 391
Db 121 CTGGCCGGAAGCCCGCGCTGCTA 144

RESULT 11
US-09-621-976-1583/c
; Sequence 1583, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621.976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 1583
; LENGTH: 455
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 143..433
; NAME/KEY: sig_peptide
; LOCATION: 143..217
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 6
; OTHER INFORMATION: seq LIGLIVAVATVHL/VI
US-09-621-976-1583

Alignment Scores:
Pred. No.: 136          Length: 455
Score: 8.00           Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.70%      Indels: 0
DB: 4                  Gaps: 0

US-09-993-966-7 (1-470) x US-09-621-976-1583 (1-455)

QY 334 ArgLeuArgGlyThrGlnaSPGLY 341
Db 332 CGACTCCGGGGAAGCTCAAGATGTT 309

RESULT 12
US-09-270-767-3943

; Sequence 3943, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3943
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-3943

Alignment Scores:
Pred. No.: 137          Length: 458
Score: 8.00           Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.70%      Indels: 0
DB: 4                  Gaps: 0

US-09-993-966-7 (1-470) x US-09-270-767-3943 (1-458)

QY 456 HisHisHisGluHisHisHis 463
Db 300 CATCACACGAGCAGCACACACAC 323

RESULT 13
US-09-270-767-19225
; Sequence 19225, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 19225
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-19225

Alignment Scores:
Pred. No.: 137          Length: 458
Score: 8.00           Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.70%      Indels: 0
DB: 4                  Gaps: 0

US-09-993-966-7 (1-470) x US-09-270-767-19225 (1-458)

QY 456 HisHisHisGluHisHisHis 463
Db 300 CATCACACGAGCAGCACACACAC 323

RESULT 14
US-09-270-767-5698/c
; Sequence 5698, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
```

SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 5698
 LENGTH: 482
 TYPE: DNA
 ORGANISM: Drosophila melanogaster
 US-09-270-767-5698

Alignment Scores:
 Pred. No.: 144 Length: 482
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.70% Indels: 0
 DB: 4 Gaps: 0

US-09-993-966-7 (1-470) x US-09-270-767-5698 (1-482)

QY 70 ValLeuArgAspThrLeuSerGlu 77
 |||||
 DB 48 GTTTAAGACACCTTAACGAA 25

RESULT 15

US-09-270-767-20980/c
 Sequence 20980, Application US/09270767
 Patent No. 6703491
 GENERAL INFORMATION:
 APPLICANT: Homburger et al.
 TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 FILE REFERENCE: File Reference: 7326-094
 CURRENT APPLICATION NUMBER: US/09/270,767
 CURRENT FILING DATE: 1999-03-17
 NUMBER OF SEQ ID NOS: 62517
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 20980
 LENGTH: 482
 TYPE: DNA
 ORGANISM: Drosophila melanogaster
 US-09-270-767-20980

Alignment Scores:
 Pred. No.: 144 Length: 482
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.70% Indels: 0
 DB: 4 Gaps: 0

US-09-993-966-7 (1-470) x US-09-270-767-20980 (1-482)

QY 70 ValLeuArgAspThrLeuSerGlu 77
 |||||
 DB 48 GTTTAAGACACCTTAACGAA 25

Search completed: December 30, 2004, 13:44:46
 Job time: 148 secs

THIS PAGE IS BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 30, 2004, 07:54:31, Search time 5643 Seconds
(without alignments)
3035.030 Million cell updates/sec

Title: US-09-993-966-7
Perfect score: 470
Sequence: 1 MGLKSKPAVCKRSPESG.....RHHHHHHHHHHHHFYOT 470

Scoring table: @E60
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Word size: 1

Total number of hits satisfying chosen parameters: 65644297

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xip
-O/cgnt 1/USPTO.spool_P/US0993966/runat 29122004.130829.9394/app query.fasta.1.647
-DB=EST -OPMT=fastlap -SUPFIX=olig.rst -MINMATCH=0 -LOOPTL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALLGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptco
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0993966 @CGN 1.1.6425 @runat 29122004.130829.9394 -NCPU=6 -ICPU=3
-NO MMAP -LARGEIOBERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6
-Fgapext=7 -Ygapop=60 -Ygapext=60 -DELOP=6 -DELEXT=7

Database: EST:*

1: gb_est1:.*
2: gb_est2:.*
3: gb_hlc:.*
4: gb_est3:.*
5: gb_est4:.*
6: gb_est5:.*
7: gb_est6:.*
8: gb_ges1:.*
9: gb_ges2:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	460	97.9	2142	3	AF289584 Homo sapi
2	295	62.8	885	9	AY412098 Homo sapi
3	270	57.4	931	5	B0653673 AGENCOURT
4	248	52.8	921	5	B0645656 AGENCOURT
5	235	50.0	895	5	B0644360 AGENCOURT
6	230	48.9	996	5	B0652087 AGENCOURT
7	227	48.3	990	5	B0064678 AGENCOURT
8	212	45.1	1008	5	B0646371 AGENCOURT
9	194	41.3	627	4	B1767278 603057995

10	187	39.8	928	5	B0644956 AGENCOURT
11	187	39.8	966	5	B0649813 AGENCOURT
12	187	38.5	953	5	B0846054 AGENCOURT
13	178	37.9	899	5	B0645507 AGENCOURT
14	169	36.0	564	4	BM711145 UI-E-DX1
15	158	33.6	622	6	CB215756 NISC np09
16	158	33.6	825	6	CB961961 AGENCOURT
17	151	32.1	1027	4	BG104777 AGENCOURT
18	150	31.9	751	4	BG820139 AGENCOURT
19	145	30.9	634	7	CN369119 170005999
20	142	30.2	429	7	CN369118 170005322
21	142	30.2	481	4	B1047069 MR3-FN020
22	136	28.9	826	6	CB961668 AGENCOURT
23	130	27.7	1059	5	BQ070932 AGENCOURT
24	104	22.1	777	4	BG542261 AGENCOURT
25	101	21.5	885	9	AY412099 Pan trogl
26	96	20.4	654	9	AG112447 Pan trogl
27	96	20.4	931	5	B0856404 AGENCOURT
28	65	13.8	395	1	A1167910 AGENCOURT
29	55	11.7	565	2	BF920913 AGENCOURT
30	49	10.4	465	7	W78547 me79606.r1
31	49	10.4	533	7	CN664615 A0815D03
32	49	10.4	540	5	BX513216 AGENCOURT
33	49	10.4	674	6	CB056689 NISC j119
34	49	10.4	677	7	CF731367 UI-M-G20
35	49	10.4	684	7	CK450648 903739 MA
36	49	10.4	724	6	CA328380 UI-M-FY0
37	49	10.4	888	9	AY412100 Mus muscu
38	49	10.4	913	5	B0895164 AGENCOURT
39	49	10.4	915	6	CB182455 AGENCOURT
40	49	10.4	934	2	BF167269 601775743
41	49	10.4	946	6	BY712252 AGENCOURT
42	49	10.4	1466	3	AK013241 Mus muscu
43	46	10.4	2471	3	AK076164 Mus muscu
44	46	9.8	555	5	BP446790 AGENCOURT
45	45	9.6	434	7	W66677 me19605.r1

ALIGNMENTS

RESULT 1	AF289584	2142 bp	mRNA	linear	HTC 01-JAN-2002
LOCUS	AF289584				
DEFINITION	Homo sapiens clone pp7246 unknown mRNA.				
ACCESSION	AF289584				
VERSION	AF289584.1	GI:18027371			
KEYWORDS	HTC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 2142) Zhang,P.P., Zhou,X.M., Jiang,H.Q., Huang,Y., Qin,W.X., Zhao,X.T., Wan,P.F. and Gu,J.R.				
TITLE	Novel human cDNA clones with function of inhibiting cancer cell growth				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 2142) Zhang,P.P., Zhou,X.M., Jiang,H.Q., Huang,Y., Qin,W.X., Zhao,X.T., Wan,P.F. and Gu,J.R.				
AUTHORS	Man,D.F. and Gu,J.R.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-JUL-2000) National Laboratory For Oncogenes & Related Genes, Shanghai Cancer Institute, 25/Jn 2200 Xie-Tu Road, Shanghai 200032, P. R. China				
FEATURES	location/Qualifiers				
source	1..2142				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="pp7246"				
	277..1689				
	/codon_start=1				
	/product="unknown"				

/protein_id="AAU55768.1"
/db_xref="GI:18027372"
/translation="MGKLSKPYDVCRRSPGSDPAVSAAMAKIEEMIGRORP
GVSGPRQLAAGTIGRSRELVVDVLRDLTSEBEDDFLEVALPEPKTDIGSGDS
KMERVSEPCPGSKQKLFELQCDVSMEDSRQEWFTLYDPNNKRVREDTLSL
HTLYEVVDSVSNHSPTSKMLRYVLTVAPDSQSRSVLVNOADLQARPRAEKPT
DLSEMEKQRAPLRFQGDRLQSGCHNRKQVDEINERNNHYLDLAGIENTSGPGPS
PSVAOKSELPPRTSNPTRSRSHPEALHHPKQGVDPASPHFLDTPIAKVSELQOR
LRGTQDSKHFVSPKQSGSVGVGHARARANKPRLGPAIPAVSPAHIAASPLLP
SLAPLGHKKHKKAKESQCGRGLQAPLASGVPVLGHEHLRELPALVTVESQKQPVQ
RHEHHHHHHHHHHYHFQY"

ORIGIN

Alignment Scores:

Pred. No.:	0	Length:	2142
Score:	460.00	Matches:	460
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.87%	Indels:	0
DB:	3	Gaps:	0

US-09-993-966-7 (1-470) x AF289584 (1-2142)

QY 11 ValCysIysArgArgGluSerProGluGlyAspSerPheAlaValIserAlaAlaTrpAla 30
Db 307 GTGTGCAAGCGCAGGAGAGCCCGAAAGGTGACAGCTTCGCGGTGAGCGCTGGGCT 366
QY 31 ArgIysGlyIleGluGluTrpIleGlyValArgGlnArgCysProGlyGlyValSerGlyPro 50
Db 367 CGGAAGGACATCGAGGAGTGTATCGGAGACAGCGCTGCCGCGGTGTCTCGGAGACC 426
QY 51 ArgGlnLeuArgLeuAlaGlyThrIleGlyValArgSerThrArgGlnLeuValGlyAspVal 70
Db 427 CGACAGCTGGCGTTGGCGGACCAATAGCCGAGACACCGGAGCTTGTTGGCGAGCTG 486
QY 71 LeuArgSerThrLeuSerGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 90
Db 487 TTGAGAGACACGCTCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 546
QY 91 ProGluSerThrArgGlyLeuGlySerGlyValArgGlnValArgValSerGlu 110
Db 547 CTGTAGAGAGCTGACGGGCTGGGCGAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAG 606
QY 111 ProCysProGlySerIleGlyValLeuLeuPheGlnGluGlnGlnGlnGlnGlnGlnGlnGln 130
Db 607 CCTGCGCCAGGCTCCAGAGAGACCTGAAGTTTAAAGCTCCAGTGGAGCTGTCCATG 666
QY 131 GlnGluAspSerArgGlnGluTrpThrPheThrLeuTyrAspPheAspAsnGlyLys 150
Db 667 GAGGAGACAGCCGCGAGAGTGAACCTTCACTGTATGACTTGAACAACAGCGCAG 726
QY 151 ValThrArgGluAspIleThrSerLeuLeuHisThrIleTyrGluValValAspSerSer 170
Db 727 GTCAACCCGAGAGGACATCACAGCTTCTGACACCACTATAGAGGTGGTGGACCTCTCT 786
QY 171 ValAsnHisSerProThrSerSerIleMetLeuArgValIlysLeuThrValAlaProAsp 190
Db 787 GTCAACCACTCCCAACATCCAGAAATGCTGGGTTAAAGCTCACCGTGCCCGCAT 846
QY 191 GlySerGlnSerIleValSerValLeuValAsnGlnAlaAspLeuGlnSerAlaArgPro 210
Db 847 GCGAGCCAGAGCAGAGAGAGGCTCTTGTCAATCAGGCTGACCTGCAAGAGCGCAAGGCC 906
QY 211 ArgAlaGluThrIleValProThrGlnAspLeuArgSerTrpGluValValGlnArgAlaPro 230
Db 907 CGAGCAAGACCAAGCCCACTGAGAGACCTGCGGCTGGGAGAGAGAGAGAGAGAGAGAG 966
QY 231 LeuArgPheGlnGluValAspSerArgLeuGlnGlnSerGlyCysTyrTrpHisCysValAsp 250
Db 967 CTCAGGTTCCAGGGTGAAGCCCGCTGAGAGAGTGTGGCTGCTACCAACATTTGGTGAT 1026
QY 251 GlnAsnIleGluValArgAsnHisTyrLeuAspLeuAlaGlyIleGluAsnTyrThrSer 270

Db 1027 GAGAACATCAGAGAGAGAAACCACTACTTATGATCTCGCCGGAGTAGAAAACTACACGTCC 1086
QY 271 GlnPheGlyProGlySerProSerValAlaGlnIleSerGluLeuProProAlaGlyThrSer 290
Db 1087 CAATTGGGCTGAGCTCCCTTCCTCGTGGCCCAAGAGCAAGACTGCCCCCGACCTCC 1146
QY 291 AspProThrArgSerArgSerHisGluProGluAlaIleHisIleProHisArgLysPro 310
Db 1147 AATCCCACTGATCTCGCTCCCATGAGCCGGAAACCATCCACATCCACAGCCGAAAGCCC 1206
QY 311 GlnGlyValAspProAlaSerPheHisPheLeuAspThrProIleAlaValSerGlu 330
Db 1207 CAAGCGTGAACCCGGCTCTCTTCACTTCTTGACACCCCAATCGCAAGGTCTCAGAG 1266
QY 331 LeuGlnGlnArgLeuArgGlyThrGlnAspGlySerTyrHisPheValAspSerProLys 350
Db 1267 CTCAGACAGAGCTCGCGGACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1326
QY 351 AlaGlnGlyLysSerValGlyValAlaGlyHisValAlaArgGlyValArgAsnLysProPro 370
Db 1327 GCCAGGAGCAAGAGTGTGGTGTGGCCACGTGGCCAGAGGGGCAAGAAACAAGCCCCCT 1386
QY 371 LeuGlyProAlaIleProAlaValSerProSerAlaHisIleValAlaSerProAlaLeu 390
Db 1387 CTGGAGACCCGCAATCCCTCGGTGTGTCCCTCCGCGCACTGGGTGCGAGCCGGCCCTC 1446
QY 391 LeuProSerLeuAlaProLeuGlnHisIleValHisIleValHisIleValHisIleValHis 410
Db 1447 CTCCTCTCTTACCCCTCGGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1506
QY 411 GlnGlyCysArgGlyLeuGlnAlaProLeuAlaSerGlyValProValLeuGlyArgGlu 430
Db 1507 CAGGCGTCCGGGGGCTCGAGAGCAACACTGGCTCAGGTGGCCCTGCTCGGGCGGAG 1566
QY 431 HisLeuArgGluLeuProAlaLeuValValTyrGluSerGlnAlaGlyGlnProValGln 450
Db 1567 CACCTGGGAGAGTGGCCGCTTGTGTGTATAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1626
QY 451 ArgHisGlnHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 470
Db 1627 AGACATGAGACACACACACACATGACATCACCATTAACACACTTCTTACAGACA 1686

RESULT 2
AY412098 885 bp DNA linear GSS 16-DEC-2003
LOCUS
DEFINITION Homo sapiens NKD1 gene, VIRUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY412098
VERSION AY412098.1 GI:39768063
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Chordata; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 885)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civeillo,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,D.J.,
Todd,M.A., Tanenbaum,D.M., Civeillo,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,D.J.,
Adams,M.D. and Cargill,M.
Interferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
TITLE
JOURNALS Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
2 (bases 1 to 885)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civeillo,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,D.J.,
Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT
This sequence was made by sequencing genomic exons and ordering
them based on alignment.


```

Db      123 CAGTCTGGCTGCTACCACTGGGTGATGAGAACTCCAGAGAAACCACTACTTA 182
Qy      261 Aspleu1a1g1y1eg1uean1y1r1Th1Ser1Gln1Phe1Gly1Ser1Pro1Ser1Phe1Val1a 280
Db      183 GATCTCCCGGAGTAGAAAATCACTACGCTCCCAATTTGGGCTGCTCCCTTCGCTGCC 242
Qy      281 G1LysSerG1LueuP1roP1oArG1Th1SerAsn1Pro1Th1rAsa1rAsa1rG1u1Pro 300
Db      243 CAGAGGTCAGAACTGCCCCCGACCTCCATCCATCCATCGATCTGCTCCCATGACCG 302
Qy      301 G1u1a1l1e1H1s1l1e1Pro1H1s1rG1y1S1rG1nG1y1Val1Asp1ro1a1Ser1Phe1Phe 320
Db      303 GAGGACATCCACATCCACACCGAAAGCCCAAGGCGGTGACCCGCGCTCTTCCACTTC 362
Qy      321 LeuAsp1Th1rPro1le1a1y1s1Val1Ser1G1u1e1u1nG1n1rG1e1u1a1rG1y1Th1rG1a1sp 340
Db      363 CTGACACACCCCAATCCGCAAGGTCTCAGAGCTCCAGCACGGCTCCGCGGACCCAGGAC 422
Qy      341 G1y1Ser1y1H1s1Phe1Val1rG1Ser1Pro1y1a1aG1nG1y1y1Ser1Val1G1y1H1s 360
Db      423 GGGAGCAGACCTTGTGAGGTCCCGAAGGCCCAAGGAGTGTGGGTGGGCTCAC 482
Qy      361 Val1a1a1rG1y1a1a1rG1a1n1y1s1Pro1leuG1y1Pro1a1a1le1Pro1a1a1Ser1Pro 380
Db      483 GTGGCCAGAGGGGCAAGAAACACCCCTCTGGGACCCGCGCATCCCTGCGGTCTCCCC 542
Qy      381 Ser1a1H1s1e1u1a1a1a1Ser1Pro1a1a1e1u1e1u1P1ro1Ser1e1u1a1P1ro1e1u1y1H1s1y1s 400
Db      543 TCCGCCACCTGGCTGCGCAGCCCGGCTCTCCCTCCCTAGACCCCTCGGACACAG 602
Qy      401 Ly1H1s1y1H1s1A1rG1a1y1s1G1u1Ser1G1nG1y1y1s1rG1y1e1u1nG1a1a1Pro1e1u 420
Db      603 AAGCACAAGACACGAGCCAGAGAGACGACGAGGCTGCGGAGCTTGCAAGCACACTG 662
Qy      421 A1a1Ser1G1y1G1y1Pro1Val1e1u1G1y1rG1y1H1s1e1u1a1rG1u1e1u1Pro1a1e1u1Val1 440
Db      663 GCCTCAAGTGGCCCTGTCTGGGGGGGAGCAGCTGCGGAGCTGCCGCGCTTGCTGTG 722
Qy      441 Ty1rG1u1Ser1G1a1a1G1y1n1Pro1Val1G1n1rG1H1s1e1u1H1s1H1s1H1G1u1H1s 460
Db      723 TATGAGAGCCAGGCGGCGGAGCCGCTCAGAGACATGACACCAACCATGAACAT 782
Qy      461 H1s1H1s1Ty1rH1s1H1s1Phe1Ty1rG1n1Th1r 470
Db      783 CACCACCATTAACCACTTCTACCAACA 812

```

RESULT 4
B0645656 921 bp mRNA linear EST 15-JUN-2002
LOCUS AGENCOURT 8355700 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6285268
DEFINITION 5', mRNA sequence.
ACCESSION B0645656
VERSION B0645656.1 GI:21769828
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens (human)
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 921)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLCM2484 row: d column: 05

```

FEATURES
Source
High quality sequence start: 16
High quality sequence stop: 685.
Location/Qualifiers
1..921
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6285268"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_100"
/note="Organ: liver; Vector: pOTB7; Site: 1: XhoI, Site: 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(C). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 5.81e-234 Length: 921
Score: 248.00 Matches: 248
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 52.77% Indels: 0
DB: Gaps: 0

US-09-993-966-7 (1-470) x B0645656 (1-921)
Qy      202 G1u1a1a1spleu1n1ser1a1a1rG1a1a1g1u1Th1r1y1Pro1Th1r1G1u1a1spleu1a1rG 221
Db      25 CAGGCTGACTTGAGAGCGGAGGCGCCGAGCAGAGACCAAGCCCATGAGACTTCGG 84
Qy      222 Ser1rG1u1y1s1y1s1G1n1rG1a1a1P1ro1e1u1a1rPhe1G1nG1y1a1Ser1a1rG1e1u1G1n 241
Db      85 AGCTGGAGAGAAAGACGACGAGCCCGCTCAGGTTCCAGGATGACAGCCGCTCGAGCAG 144
Qy      242 Ser1y1Cy1s1Ty1rH1s1H1s1Cy1s1Val1Asp1G1u1a1n1e1G1u1a1rG1a1n1s1Ty1rLeu1a1sp 261
Db      145 TCTGGCTGCTACACCACTTCGTAAGATGAGAAATGAGAGGAGAAACCACTTAGAT 204
Qy      262 Leu1a1G1y1l1e1G1u1a1n1y1r1Th1r1Ser1G1n1Phe1G1y1Pro1G1y1Ser1Pro1a1a1a1G1n 281
Db      205 CTCGCGGAGTAGAAATCACTACGTCCTCAATTTGGGCTGCTGCTCCCTTCGAGCCAG 264
Qy      282 Ly1Ser1G1u1e1u1P1ro1P1oArG1Th1r1Ser1a1n1Pro1Th1r1rG1Ser1a1rG1e1rH1s1G1u1Pro1G1u 301
Db      265 AAGTCAGAACTGCCCCCGCCACCTCCATCTCCACTGATCTGCTCCCATGAGCCGGA 324
Qy      302 A1a1l1e1H1s1l1e1Pro1H1s1rG1y1S1rG1nG1y1Val1Asp1ro1a1Ser1Phe1H1s1Phe1e1u 321
Db      325 GCCATCCACATCCACACCCAAAGCCCAAGGCGGTGAGCCGCGCTCTCCACTTCCTT 384
Qy      322 Asp1Th1rPro1le1a1y1s1Val1Ser1G1u1e1u1nG1n1rG1e1u1a1rG1y1Th1rG1n1a1sP1y 341
Db      385 GACACCCCAATCCGCAAGGTCTCAGAGCTCAGACACGCTCCGAGGACCCAGAGCGG 444
Qy      342 Ser1y1H1s1Phe1Val1rG1Ser1Pro1y1a1aG1nG1y1y1Ser1Val1G1y1H1s1Val1 361
Db      445 AGCAGACACTTGTGAGGTCTCCCAAGGCCCAAGGCAAGGTGTGGGTGTGGGCACTG 504
Qy      362 A1a1rG1y1a1a1rG1a1n1y1s1Pro1leuG1y1Pro1a1a1le1Pro1a1a1Ser1Pro1Ser 381
Db      505 GCCAGAGGGGCAAGAAACAGCCCTCTGCGGACCCGCTGCGGTGTGCTCCCTCC 564
Qy      382 A1a1H1s1e1u1a1a1a1Ser1Pro1a1a1e1u1e1u1P1ro1Ser1e1u1a1P1ro1e1u1y1H1s1y1s 401
Db      565 GCCCAGCTGCTGCGCAGCCCGGCTCTCCCTCCCTTACCCCTCGGAGCACAAAG 624
Qy      402 H1s1y1H1s1A1rG1a1y1s1G1u1Ser1G1nG1y1y1s1rG1y1e1u1nG1a1a1Pro1e1u1a1 421

```


Db 625 CACAGACCGGACCGGAGAGAGCGGAGCGGCGGCGCTCGACGACCACTGCCC 684

QY 422 SerGlyProValLeuGlyArgGluHisLeuArgGluLeuProAlaLeuValTyr 441

Db 685 TCAGGTGGCCCTCTCTCGGGCGGAGACCTCGGAGAGCTGCGGCTTGTTGTTAT 744

QY 442 GluSerGlnAlaGlyGlnProVal 449

Db 745 GAGAGCCAGGCGGAGCGGCGGCTC 768

RESULT 5
B0644360 895 bp mRNA linear EST 15-JUN-2002
LOCUS AGENCOURT_8351368 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6286917
DEFINITION 5', mRNA sequence.
ACCESSION B0644360
VERSION B0644360.1 GI:21768532
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 895)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Straube, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LNCM2488 row: h column: 22
High quality sequence stop: 622.
Location/Qualifiers
1. 895
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6286917"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH MGC 100"
/note="Organ: liver; Vector: pOT7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming; Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH-MGC
Library."

ORIGIN

Alignment Scores:

Prod. No.: 3 95e-221 Length: 895
Score: 235.00 Matches: 235
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 50.00% Indels: 0
DB: 5 Gaps: 0

US-09-993-966-7 (1-470) x B0644360 (1-895)

QY 202 GlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrArgProThrGluAspLeuArg 221

Db 7 CAGGCGACCTGACAGCGGACGCGGCGGAGAGCGGAGCGGAGCGGAGCGGAGCGG 66

QY 222 SetTrpGlnLysLysGlnArgAlaProLeuArgPheGlnLysPheSerArgLeuGln 241

Db 67 AGCTGGAGAGAGAGAGAGAGCGGCGGCGGCGGCTCAGGTTCCAGGTTACAGCGGCGGAGCAG 126

QY 242 SerGlyCysTrpHisHisCysValAspGluAsnIleGluArgArgAsnHisTrpLeuAsp 261

Db 127 TCCTGGCTGACACACCACTTCGTAAGATGAGATCGAGAGGAGAAACCACTACTTAGAT 186

QY 262 LeuAlaGlyTyrIleGluAsnTrpHisArgGlnPheGlyProGlySerProSerValAlaGln 281

Db 187 CTCGCGGAGATGAAACCTACAGCTCCCAATTTGGCTTGGCTTCTCTCTCTGAGCCAG 246

QY 282 LysSerGluLeuProProArgThrSerAsnProThrArgSerArgSerHisGluProGlu 301

Db 247 AAGTCAGAACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 306

QY 302 AlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPheLeu 321

Db 307 GCCATCCATCCACACCGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 366

QY 322 AspThrProIleAlaLysValSerGluLeuGlnArgLeuArgGlyThrGlnAspGly 341

Db 367 GACACCCCAATCCGCAAGGCTTCAAGCTTCACAGCTTCGCGGCGGCGGCGGCGGCGG 426

QY 342 SerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValGlyHisVal 361

Db 427 AGCAGACACTTTGTAGGTTCCCGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 486

QY 362 AlaArgGlyAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerProSer 381

Db 487 GCCAGAGGGGCAAGAAACAAGCCCTCTGTGAGCCGCGCATCTCCGCGGTGCTCCCTCC 546

QY 382 AlaHisLeuAlaLysProAlaLeuLeuProSerLeuAlaProLeuGlyHisLysLys 401

Db 547 GCCCAGCTGCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 606

QY 402 HisLysHisArgAlaLysGlnSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeuAla 421

Db 607 CACAGACCGGACCGGAGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 666

QY 422 SerGlyProValLeuGlyArgGluHisLeuArgGluLeuPro 436

Db 667 TCAGGTGGCCCTCTCTCGGGCGGAGACCTCGGAGAGCTGCCC 711

RESULT 6
B0652087 996 bp mRNA linear EST 15-JUN-2002
LOCUS AGENCOURT_8207495 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6283110
DEFINITION 5', mRNA sequence.
ACCESSION B0652087
VERSION B0652087.1 GI:21776259
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 996)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Straube, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LNCM2478 row: j column: 07
High quality sequence stop: 604.
Location/Qualifiers
1. 996
/organism="Homo sapiens"

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_image="6283110"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_100"
/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGGACGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

ORIGIN

Alignment Scores:

Score: 3.8e-216 Length: 996
230.00 Matches: 230
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 48.94% Indels: 0
DB: 5 Gaps: 0

US-09-993-966-7 (1-470) x BQ652087 (1-996)

QY 203 AAlaAspleuGInSerAlaArgProAArgAlaGluThrIysProThrgluAspleuAysSer 222
DB 9 GCTGACCTGCAGAGCGCAGAGCCCGGAGAGACCAAGCCACCTGAGGACCTCGAGAGC 68
QY 223 TTrgIuIySylsGInAArgAlaProLeuAArgPheGlnGlyAspSerArgleuGluGlnSer 242
DB 69 TGGGAGAGAGAGCGAGCGAGCCCGCTCAGGTCAGGCTGACAGCCCGCTGGAGACAGTCT 128
QY 243 GilyCyetyrHisHisCysValAspGluAsnIleGluArgArgAsnHisIleTyLeuAspleu 262
DB 129 GGCGGTACACCACTTCGTAGATGAGAACATCGAGAGGAGAAACCACTTATGATCTC 188
QY 263 AlAGlyIleGluAenTyrrThrSerGlnPheGlyProGlySerProSerValAlaGlnIly 282
DB 189 GCGGGGTAGAAAACCTACAGCTCCCAATTGGGGCTGCTCCCTCCCTGCGCCAGAG 248
QY 283 SerGluLeuProAArgThrSerAsnProThraSerArgSerHisGluProGluAla 302
DB 249 TCGAATCTGGCCCCCGACCTCCAAATCCACTGATCTCCTCCATGACCGGAAAGCC 308
QY 303 ILeHisIleProHisArgIysProGlnGlyValAspProAlaSerPheHisPheLeuAsp 322
DB 309 ATCCACATCCCAACCGAAGCCCGCAAGCGGTGACCGGCTCTTCCACTTCTTGAC 368
QY 323 ThrProIleAlaIysValSerGluLeuGlnGlnArgleuAArgIlyThrGlnAspGlySer 342
DB 369 ACCCAATCCCGCAAGGTCTCAGAGCTCCAGCAAGCGCTCCGCGGACACCCAGAGAGCG 428
QY 343 LysHisPheValArgSerProIySerValAlaGlnGlySerValGlyValAla 362
DB 429 AAGGACTTGTGAGGTCCCGAAGGCCCAAGGAGTGGTGGTGGGCAAGTGGCC 488
QY 363 ArgGlyIlaArgAenTyrrProLeuGlyProAlaIleProAlaValSerProSerAla 382
DB 489 AGAGGGGCAAGAAACAAGCCCTCTGGAGCCGCGCATCCCTGGGTGTCCCTCCGCC 548
QY 383 HisLeuAlaIaSerProAlaLeuLeuProSerleuAlaProLeuGlyHisIlyValHis 402
DB 549 CACCTGGCTGCAGCCCGGCTCTCTCCCTCCCTAGCCCTCCGAGCAACAAGAGAC 608
QY 403 LysHisArgIaIySylsGInAArgAlaProLeuAArgPheGlnGlyAspSerArgleu 422
DB 609 AAGGACGAGCGAAGAGCGCAGCAGGCTGCGGGGCTGCGAGGACCACTGCGCTCA 668
QY 423 GilyGlyProValleuGlyArgGlnHisIleu 432

DB 669 GGTGGCCCTGTCTGGGGCGGAGACACTGT 698

RESULT 7
BQ064678 990 bp mRNA linear EST 02-APR-2002
LOCUS
DEFINITION
AGENCOURT 6853565 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5926427
5', mRNA sequence.
BQ064678
BQ064678.1 GI:19893537

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
NIH-MGC <http://mgc.nci.nih.gov/>,
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strauberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
[found through the I.M.A.G.E. Consortium/LNL at:](http://image.jnl.gov)
<http://image.jnl.gov>
Plate: LUCM2099 Row: 1 Column: 12
High quality sequence stop: 670.
Location/Qualifiers

FEATURES
source

1..990
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_image="5926427"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_99"
/note="Organ: lymph. Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGGACGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

ORIGIN

Alignment Scores:

Score: 3.48e-213 Length: 990
227.00 Matches: 227
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 48.30% Indels: 0
DB: 5 Gaps: 0

US-09-993-966-7 (1-470) x BQ064678 (1-990)

QY 201 AAsnGlnIlaAspleuGInSerAlaArgProAArgAlaGluThrIysProThrgluAspleu 220
DB 3 AATAGGCTGACTGACGAGAGCGGAGCCCGGAGAGACCAAGCCACCTGAGGACTTG 62
QY 221 ArgSerTrpGluIySylsGInAArgAlaProLeuAArgPheGlnGlyAspSerArgleu 240
DB 63 CGGAGCTGGGAGAGAGAGCGAGCGCCGCTCAGGTTCCAGGGTGAACACCGCTTGAG 122
QY 241 GlnSerGlyCyetyrHisHisCysValAspGluAsnIleGluArgArgAsnHisIleTyLeu 260
DB 123 CAGCTGGCTGTACCACTTGGTGTAGATGAGAACATCGAGAGGAGAAACCACTACTTA 182
QY 261 AspleuAlaGlyIleGluAenTyrrThrSerGlnPheGlyProGlySerProSerValAla 280

Db 183 GATCGCGCGGATAGAACTACACGTCCCAATTTGGGCTCCCTCCGTGGCC 242
 QY 281 GILYsSerGluLeuProProAqThrSerAnProThraSerAqSerHieGluPro 300
 Db 243 CAGAGTCAGAACATGCCCCCCCCGACCTCCCAATCCCACTGATCTCGCTCCATGAGCCG 302
 QY 301 GUAAlIeHieIlePheProHieAqGlyeProGlnGlyValAspProAlaSerPheHiePhe 320
 Db 303 GAAGCCATCCACATCCACACCGAAAGCCCCCAAGCCGTGAGCCGCGCTCTTCACCTTC 362
 QY 321 LeuAspThrProIleAlaValSerGluLeuGlnGlnAqGlyeValThraAsp 340
 Db 363 CTGGACACCCCAATCCGCAAGGTCTCAGAGTCCAGCAACGCTCCGGGGACCCAGAGAC 422
 QY 341 GILYsSerIlePheValAqSerProGlyAlaGlnGlySerValGlyValGlyHie 360
 Db 423 GGGAGCAGACCTTTGTGAGGTCCCCCAAGGCCAGAGCCAGAGTGTGGGTGGGCAC 482
 QY 361 ValAlaArgGlyAlaAqGlyeProProLeuGlyProAlaIleProAlaValSerPro 380
 Db 483 GTGGCCAGAGGGGCAAGAAACAAGCCCCCTGTGGAGCCCGCATCTCGCGGTGTCCCC 542
 QY 381 SerAlaHieLeuAlaAqSerProAlaLeuLeuProSerLeuAlaProLeuGlyHie 400
 Db 543 TCGGCCACCTGGTGTGCACGCCGCCCTCTCTCCCTAGGCCCTCCGGGACAG 602
 QY 401 LyeHieIleValArgAlaValGlySerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
 Db 603 AAGCACAAGACCGAGCCAGAGAGCAGACAGGGGTGCGGGGCTGACAGGCACCACTG 662
 QY 421 AlaSerGlyLeuProValLeu 427
 Db 663 GCCTCAGGTGACCTGTCTG 683

RESULT 8
 LOCUS B0646371 1008 bp mRNA linear EST 15-JUN-2002
 DEFINITION AGENCOURT_8493065 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6300188
 5', mRNA sequence.
 ACCESSION B0646371
 VERSION B0646371.1 GI:21770543
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1008)
 NIH-MGC http://mgs.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strauberg, Ph.D.
 Email: c9abers-remail.nih.gov
 Tissue Procurement: CGAP (Stanford)
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLCM2514 row: a column: 21
 High quality sequence start: 17
 High quality sequence stop: 553.
 Location/Qualifiers
 1..1008
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6300188"
 /tissue_type="hepatocellular carcinoma, cell_line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 100"
 /note="Organ: liver; Vector: pOT7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned

FEATURES

source
 1..1008
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6300188"
 /tissue_type="hepatocellular carcinoma, cell_line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 100"
 /note="Organ: liver; Vector: pOT7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned

info. EcoRI/XhoI sites using the following 5' adaptor:
 GGCAAGAG(G). Size selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the Laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC
 Library."

ORIGIN

Alignment Scores:

Pred. No.:	2,34e-198	Length:	1008
Score:	212.00	Matches:	225
Percent Similarity:	99.56%	Conservative:	0
Best Local Similarity:	99.56%	Mismatches:	0
Query Match:	45.11%	Indels:	1
DB:	5	Gaps:	0

US-09-993-966-7 (1-470) x B0646371 (1-1008)

QY 202 GILAlaAspLeuGlnSerAlaAqProAqAlaGluThraProThraGluAspLeuAq 221
 Db 25 CAGGCTGACCTGACAGAGCCGCAAGCCCGAGCAGAGCCAAAGCCACTGAGGACCTCGG 84
 QY 222 SerTrpGluValLysGlnAqAlaProLeuAqPheGlnGlyAspSerAqLeuGln 241
 Db 85 AGCTGGAG 144
 QY 242 SerGlyCysTrpHieHieCysValAspGluAsnIleGluAqAqAqHieIleValLeuAsp 261
 Db 145 TGTGGCTGTACCAACCATTCGTATGATGAGAACATGAGAGAGAGAGAGAGAGAGAGAG 204
 QY 262 LeuAlaGlyIleGluAqHieIleThraSerGlnPheGlyProGlySerProSerValAlaGln 281
 Db 205 CTCGCCGAGATAGAACTACACGTCCCAATTTGGGCTGCTCCCTCCGTGGCCAG 264
 QY 282 LyeSerGluLeuProProAqThrSerAnProThraSerAqSerHieGluProGlu 301
 Db 265 AAGTCAGACTGCCCCCCCCGACCTCCATATCCACTGATCTGCTCCATGAGCCGAA 324
 QY 302 AlaIleHieIleProHieAqGlyeProGlnGlyValAspProAlaSerPheHiePhe 321
 Db 325 GCATCCACATCCCAACCGAGAGGCCCAAGGGGTGAGCCGGGCTCTTCACTTCTT 384
 QY 322 AspThrProIleAlaValSerGluLeuGlnGlnAqGlyeValAspProAlaSerPhe 341
 Db 385 GACACCCCAATCCCAAGGTCTCAGAGCTCCAGCAACGGCTCCGGGACACCCAGAGCGG 444
 QY 342 SerIleHiePheValAqSerProGlyAlaGlnGlyLysSerValGlyValGlyHieVal 361
 Db 445 AGCAGACATTTGTGAGGTCCCAAGAGGCCAGAGGAGAGAGAGAGAGAGAGAGAGAG 504
 QY 362 AlaArgGlyAlaAqGlyeProProLeuGlyProAlaIle-ProAlaValSerProSe 381
 Db 505 GCGAGAGGGGCAAGAAACAAGCCCTCTTGAGACCGGCATCCCTCGGTGTCCCTC 564
 QY 381 ValAlaHieLeuAlaAqSerProAlaLeuLeuProSerLeuAlaProLeuGlyHieVal 401
 Db 565 CGGCCACCTGTGCTGAG 624
 QY 401 SHIeIleValAqAlaValSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeuAl 421
 Db 625 GCAAGAGACCGAG 684
 QY 421 aSerGlyGlyProVal 426
 Db 685 CTCAGGTGCGCTGT 700

RESULT 9
 LOCUS B1767278 627 bp mRNA linear EST 25-SEP-2001
 DEFINITION 603057995F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5207614 5',
 mRNA sequence.
 ACCESSION B1767278

VERSION B167278.1 GI:15758869
KEYWORDS EST. Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 627)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM1520 row: n column: 23
High quality sequence stop: 580.
Location/Qualifiers
1..627
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5207614"
/lab_host="DH10B"
/clone_1lb="NIH MGC 122"
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleens. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH_MGC Library."
ORIGIN
Alignment Scores:
Pred. No.: 9.13e-181 Length: 627
Score: 194.00 Matches: 194
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 41.28% Indels: 0
DB: 4 Gaps: 0
US-09-993-966-7 (1-470) x B167278 (1-627)
QY 183 VallysleuthrVala1aProaspGlySerInserlyArServalIleuValaIn 202
Db 2 GTAAAGCTCAACGCGCCCGGATGGCAGCAGAGCAAGAGAGCTCTTGTCATCAG 61
QY 203 ALAaspleuInserAlaArgProArgAlaGluThrlyPProThrGluAspleuArgSer 222
Db 62 GCTGACCTGCAGACGCGAGGCCCGGAGCAGAGCAAGCCCACTGAGGACCTCGGAGGC 121
QY 223 TrpGluyslysgInatrgAlaProleuArpPheGlnIyAspSerArgyleGluGlnSer 242
Db 122 TGGAGAGAGAGCGAGCGAGCCCGCTCAGGTTCCAGGGTGACACCGCTGGAGCAGCT 181
QY 243 GlyCySTyrTrhIshIscYsValAspGluAsnIleGluArgArGAsnHsIYrLeuAspleu 262
Db 182 GGCTGCTACACCACTTGCTGATGAGAACATCGAGAGGAGAAACCACTACTTGATCTC 241
QY 263 ALAglY11leGluArYrThrSerGlnPheGlyProGlySerProSerValAlaGlnly 282
Db 242 GCCCGGATAGAAACCTACACGCTCCCAATTTGGGCTGCTCCCTTCGTCGAGAG 301
QY 283 SerGluuPProPArGThrSerAsnProThrArgerArGSerHsIgluProGluA 302

Db 302 TCAGAACTGCCCGCCGACCTCCAACTCCATGATCGCTCCATGAGCGGAGACC 361
QY 303 ILeHsIleProHIsaGlySProGlnIyValAspProAlaSerPheHIsPheLeuAsp 322
Db 362 ATCCACATTCACCAACCGAAGGCCCGGAGGCGTGAGCCCGGCTCTTCACCTTCCTTGAC 421
QY 323 ThProIleAlaYsValSerGluLeuGlnIyArGleuArGlyThrGlnAspGlySer 342
Db 422 ACCCAATCCCGCAGAGCTCTAGAGCTCCAGCAAGCTCCGGGCACTCCAGAGCGGAGC 481
QY 343 LyshIshPheValArGSerProLyAlaGlnIyYsSerValGlyValGlyHsValAla 362
Db 482 AAGACTTTGTGAGGTCCCGCCAGAGCGCCAGGCAAGAGTGTGGGTGGCCACGTGGCC 541
QY 363 ArgGlyAlaArgAsnlySProProleuGlyProAlaIlePro 376
Db 542 AGAGGGGCAAGAAACAGCCCGCTGGGAGCCGCGCATCC 583
RESULT 10
LOCUS B0644956 928 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8511562 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6297032
5', mRNA sequence.
ACCESSION B0644956
VERSION B0644956.1 GI:21769128
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 928)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM2505 row: n column: 09
High quality sequence start: 17
High quality sequence stop: 551.
Location/Qualifiers
1..928
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6297032"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH MGC 100"
/note="Organ: liver; Vector: pOT87; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
ORIGIN
Alignment Scores:
Pred. No.: 1.08e-173 Length: 928
Score: 187.00 Matches: 236
Percent Similarity: 99.16% Conservative: 0
Best Local Similarity: 99.16% Mismatches: 1
Query Match: 39.79% Indels: 2
DB: 5 Gaps: 0

US-09-993-966-7 (1-470) x B0644956 (1-928)

```

QY 202 GlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIysProThrGluAspLeuArg 221
Db 21 CAGGCTGACCTGCAAGAGCGGAGGCCCGGAGCAAGACCAAGCCACTGAGGACTTCGG 80
QY 222 SerTTGGLuLysLysGlnArgAlaProLeuArgPheGlnIysAspSerArgLeuGluGln 241
Db 81 AGCTGGAGAGAGAGAGCGAGCGCCCGCTCAGGTTCCAGGGGTGACACCGCGCTGGAGCAG 140
QY 242 SerGlyCySerTrpHisHisCysValAspGluAsnIleGluArgArgAsnHisIleLeuAsp 261
Db 141 TCTGGCTGCTACCACTTCCTGCTAGATGAGACATCGAGAGAGAAACCACTACTTACAT 200
QY 262 LeuAlaGlyIleGluAsnTrpTrpSerGlnPheGlyProGlySerProSerValAlaGln 281
Db 201 CTCGGCGGAGTAGAAACTACAGCTCCCAATTTGGGCTGGCTCCCTTCGGTGGCCAG 260
QY 282 LysSerGluLeuProProArgTrpSerAsnProThrArgSerArgSerHisGluProGlu 301
Db 261 AAGTCAGAACATGGCCCGCCCGCACCTCCATCCCATCTGCTCCCTCCATGAGCCGGA 320
QY 302 AlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPheLeu 321
Db 321 GCCATCCATCCCAACCCAGAACCCCAAGGCGTGAACCGGCTTCCTTCCTTCACTTCTT 380
QY 322 AspThrProIleAlaLysValSerGluLeuGlnGlnArgLeuArgGlyThrGlnAspGly 341
Db 381 GACACCCCAATCCCAAGAGTCTCAGAGCTCCAGAACGGGCTCGGGGACACCGAGAGGG 440
QY 342 SerIleHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValHisVal 361
Db 441 AGCAACACATTTGTAGGTCCCGCCAGAGCCCAAGGCAAGAGTGGGTGGGGCCAGCGT 500
QY 362 AlaArgGlyValAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerProSer 381
Db 501 GCCAGAGGGGCAAGAAACAGAGCCCGCTCTGGAGACCCCGCATCTCTGGAGTCCCTCC 560
QY 382 AlaHisLeuAlaAlaSerProAla-LeuLeuProSerLeuAlaProLeuGlyHisLysLys 401
Db 561 GCCACCTGGCTCGAGCCGAGN-CTCTCTCCCTCTCCCTAGCCGCCCTCGGGCAAGAA 619
QY 401 SerIleHisPheValArgLysGlnSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeuAl 421
Db 620 GCACAGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 679
QY 421 AspArgGlyProValLeuGlyValArgGlyHisLysLeuArgGluLeuProAlaLeu 438
Db 680 CTCAGGTGGCTGTCTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 731

```

```

RESULT 11
B0649813 966 bp mRNA linear EST 15-JUL-2002
LOCUS B0649813
DEFINITION AGENCOURT_9302499 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6271026
ACCESSION B0649813
VERSION B0649813.1 GI:21773965
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 966)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rcmail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov

Plate: LNCM2447 row: b column: 19

High quality sequence stop: 583.

Location/Qualifiers

FEATURES

source

```

1. 966
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6271026"
/issue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_100"
/note="Organ: liver; Vector: pOT7; Site_1: XhoI, Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Size-selected >500bp for average insert size
1 kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

```

ORIGIN

Alignment Scores:

```

Pred. No.: 1,13e-173 Length: 966
Score: 187.00 Matches: 228
Percent Similarity: 99.13% Conservative: 0
Best Local Similarity: 99.13% Mismatches: 1
Query Match: 39.79% Indels: 2
DB: 5 Gaps: 0

```

US-09-993-966-7 (1-470) x B0649813 (1-966)

```

QY 202 GlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrIysProThrGluAspLeuArg 221
Db 5 CAGGCTGACCTGCAAGAGCGGAGGCCCGGAGCAAGACCAAGCCACTGAGGACTTCGG 64
QY 222 SerTTGGLuLysLysGlnArgAlaProLeuArgPheGlnIysAspSerArgLeuGluGln 241
Db 65 AGCTGGAGAGAGAGAGCGAGCGCCCGCTCAGGTTCCAGGGTGAAGCGCGCTGGAGCAG 124
QY 242 SerGlyCySerTrpHisHisCysValAspGluAsnIleGluArgArgAsnHisIleLeuAsp 261
Db 125 TCTGGCTGCTACCACTTCCTGCTAGATGAGACATCGAGAGGAGAAACCACTTACAT 184
QY 262 LeuAlaGlyIleGluAsnTrpTrpSerGlnPheGlyProGlySerProSerValAlaGln 281
Db 185 CTCGGCGGAGTAGAAACTACAGCTCCCAATTTGGGCTGGCTCCCTTCGGTGGCCAG 244
QY 282 LysSerGluLeuProProArgTrpSerAsnProThrArgSerArgSerHisGluProGlu 301
Db 245 AAGTCAGAACATCCCGCCCGCACCTCCCAATCCCACTGACTGCTCCCATGAGCCGGA 304
QY 302 AlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPheLeu 321
Db 305 GCCATCCATCCCAACCCAGAACCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 364
QY 322 AspThrProIleAlaLysValSerGluLeuGlnGlnArgLeuArgGlyThrGlnAspGly 341
Db 365 GACACCCCAATCCCAAGAGTCTCAGAGCTCCACCAACGAGTCCGGGACACCGAGAGGG 424
QY 342 SerIleHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValHisVal 361
Db 425 AGCAAGACATTTGTAGGTCCCGCAAGGCGGAGGAGAGGTGGGTGGGCGACAGTGG 484
QY 362 AlaArgGlyValAlaArgAsnLysProProLeuGlyProAlaIleProAlaValSerProSer 381
Db 485 GCCAGAGGGGCAAGAAACAGAGCCCGCTCGGGAGCCCGCATCTCTGGGTGGTCCCTCC 544
QY 382 AlaHisLeuAlaAlaSerProAla-LeuLeuProSerLeuAlaProLeuGlyHisLysLys 401

```

```

Db      545 GCCACCTGCGCTGCAGCCGCGG-CCTCTCCCTCCCTAGCCCCCTCGGGCACAAGAA 603
Qy      401 ghtlyvhiarxalalysegiuserglnnglycyarqlylenuinlaproleua1 421
Db      604 GCACAAACACCGAGCCAGAGAGCCAGAGGCTGCGGGCTGAGGACCACTGCG 663
Qy      421 aserglyglyProvalleuglyarqlyu 430
Db      664 CTCAGGTGCGCTCTCTCGGGGCGCGAG 691

RESULT 12
BU846054      953 bp      mRNA      linear      EST 16-OCT-2002
LOCUS      BU846054
DEFINITION      AGENCOURT 10413301 NIH_MGC 109 Homo sapiens cDNA clone
IMAGE:6580002 5', mRNA sequence.
ACCESSION      BU846054
VERSION      BU846054.1 GI:24030759
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 953)
AUTHORS      NIH-MGC http://mgs.nci.nih.gov/.
TITLES      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM2784 row: d column: 18
High quality sequence stop: 627.
Location/Qualifiers
1. 953
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6580002"
/tissue_type="teratocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_109"
/note="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:
XhoI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.:      9.41e-168      Length:      953
Score:      181.00      Matches:      181
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 38.51%      Indels: 0
DB:      5      Gaps: 0

US-09-993-966-7 (1-470) x BU846054 (1-953)
Qy      1 MetGlylyseurhiiserlyProvalalavalCylysarqargliuserProglugly 20
Db      224 ATGGGAAATTCTCACTCAAGCGCGCGCTGTGCAAGCGCAGAGAGCCGGAAAGT 283
Qy      21 AspSerPheAlaValSerAlaAlaTTPAlaArqlyseglyllegluglTTPilleglyArg 40
Db      284 GACAGCTTCCTCGTGCAGCGCTGCTCGAAGGCGCATCGAGAGTGGATCGGAGA 343

```

```

Qy      41 GlnArgCyseProgluglyValiserglyProArqGlnleuArqheua1aglyThrilegly 60
Db      344 CAGCGTCCCGCGCGCGGTGTCTCGGAGACCCGACAGCTGGCTTGGCGGCACCATAGGC 403
Qy      61 ArgSerThrArgGlnleuValGlyAspValleuArqAspThrleuserglnuglu 80
Db      404 CGAAGCACC CGGAGCTCGTGGCGAGCGTGTGAGAGACACGCTCAGCGAGGAAGAGAG 463
Qy      81 AspAspPheArgleuGlnValAlaleuProProGlnleuThrArqGlyleuGlysergly 100
Db      464 GACGACTTTCGGGTGAGAGGCGCTGCTCTGAGAGAGCTAGCGGCTGGCGAGCGGA 523
Qy      101 AsgGlnlyseMetGlnArqValiserGluProCyseProglyserlysllyGlnleu 120
Db      524 GATGAGAGAGAGATGAGAGAGTGAAGGAGACCTGCCAGGCTCCAGAGAGAGCTGAG 583
Qy      121 PheGlnGlnleuGlnCyseAspValiserMetGlnGlnAspSerArglnuglTTPThrPhe 140
Db      584 TTGAAGAGCTCCAGTGCAGCTGTCCATGAGAGAGACCGGCGAGTGCACCTTC 643
Qy      141 ThrleuThrAspPheAspAspAspGlylyValThrArgGlnAspThrleuSerleu 160
Db      644 ACCCTGATACCTTGAACAACGCGACAGTCACCGAGAGACATCACAGCTTCTG 703
Qy      161 HisThrIleTyrcGlnValAlaAspSerSerValAsnHisSerProThrSerSerlyMet 180
Db      704 CACACATCTATAGAGTGTGAGCTCTCTGTCAACACATCCCAATCATCAGACATG 763
Qy      181 leu 181
Db      764 CTG 766

RESULT 13
B0645507      899 bp      mRNA      linear      EST 15-JUL-2002
LOCUS      B0645507
DEFINITION      AGENCOURT 8299367 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6271502
IMAGE:6271502 5', mRNA sequence.
ACCESSION      B0645507
VERSION      B0645507.1 GI:21769679
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 899)
AUTHORS      NIH-MGC http://mgs.nci.nih.gov/.
TITLES      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM2448 row: f column: 15
High quality sequence stop: 600.
Location/Qualifiers
1. 899
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6271502"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_100"
/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size

```

ORIGIN

1 kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA Synthesis Kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC library."

Alignment Scores:
 Pred. No.: 8.2e-165 Length: 899
 Score: 178.00 Matches: 204
 Percent Similarity: 99.03% Conservative: 0
 Best Local Similarity: 99.03% Mismatches: 0
 Query Match: 37.87% Indels: 2
 DB: Gaps: 0

US-09-993-966-7 (1-470) x B0645507 (1-899)

Qy 202 GlnAlaApLeuGlnSerAlaArgProArgAlaGlnThrLysProThrGlnAlaApLeuArg 221
 Db 5 CAGGCTACCTGACAGCGGAGGCGCCGAGACAGACCAAGCCACTGAGGACCTGGCG 64
 Qy 222 SerTTPGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGln 241
 Db 65 AGCTGGAGAAAGAAAGCAGCGCGCCGCTCAGGTTCCAGGGTGAACAGCCGCTGAGACAG 124
 Qy 242 SerGlyCySerGlyrHisHisGlyValAspGluAsnIleGluArgArgAsnHisGlyrLeuAsp 261
 Db 125 TCTGGCTGCTACCACTTCGTAGATGAGAACATTCAGAGAGAAACCACTACTTGAT 184
 Qy 262 LeuAlaGlyIleGluAsnGlyrThrSerGlnPheGlyProGlySerProSerValAlaGln 281
 Db 185 CTGGCGGGATAGAAACTACACCTCCCAATTTGGCTGCTCTCCCTCCGTCGCGCAG 244
 Qy 282 LysSerGluLeuProArgThrSerAspProThrArgSerArgSerHisGluProGlu 301
 Db 245 AAGTCAGAACTGCGCCCGCCGACCTCCATCCCATCTGATCTGCTCCCATGAGCCGGA 304
 Qy 302 AlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPheLeu 321
 Db 305 GCATCCACATCCACACCCGAAAGCCCAAGCGCTGAGACCGGCTCTCCCTCACTTCCTT 364
 Qy 322 AspThrProIleAlaLysValSerGluLeuGlnIleGlnIleGlnIleGlnIleGlnIle 341
 Db 365 GACACCCCAATCGCAAGGTCCTCAGACCTCCAGCAAGCGCTCCGCGGACCCAGAGCGG 424
 Qy 342 SerLysHisPheValArgSerProLysAlaGlnGlyLysSer-ValGlyValGlyHisVal 361
 Db 425 AGCAGACACTTTGTGAGCTCCCAAGGCCAGGCGCAAGAGTGCTGGGTGGGCCCACT 484
 Qy 361 AlaArgGlyValArgAsnLysProProLeuGlyProAlaIle-ProAlaValSerProS 381
 Db 485 GGCAGAGAGGGGCAAGAAAGAGCCCTCTGGAGACCGCCATCCCTCGGGTGTCCCTCT 544
 Qy 381 erLalaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisIle 401
 Db 545 CCGGCCACCTGGCGGCGGCGGCTCTCTCCCTCTTAAGCCCTCGGCGCAAGA 604
 Qy 401 YSHIsLysHisArg 405
 Db 605 AGCACAAGCACCGG 618
 RESULT 14 564 bp mRNA linear EST 28-FEB-2002
 LOCUS BM711145
 DEFINITION UI-E-DX1-agv-m-24-0-UI.r1 UI-E-DX1 Homo sapiens cDNA clone
 ACCESSION BM711145 GI:19024403
 VERSION BM711145.1
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 564)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548

COMMENT

Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: Bento-soares@uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..564
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-DX1-agv-m-24-0-UI"
 /tissue_type="fetal eyes"
 /dev_stage="fetal"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_1lb="UI-E-DX1"
 /note="Organ: eye; Vector: pTRT3-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-E-DX1 is a normalized cDNA library containing the
 following tissue(s): fetal eyes. The library was
 constructed according to Bonaldo, Lennon and Soares,
 Genome Research, 6:791-806, 1996. First strand cDNA,
 synthesis was primed with an oligo-dT primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I
 adaptor, digested with Not I, and cloned directionally
 into pTRT3-Pac vector. The oligonucleotide used to prime
 the synthesis of first-strand cDNA contains a library tag
 sequence that is located between the Not I site and the
 (dT)18 tail. The sequence tag for this library is
 AGAATCAGA. This library was created for the program, Gene
 Discovery in the Visual System, supported by National Eye
 Institute (NEI)."

FEATURES

source

ORIGIN

Alignment Scores:
 Pred. No.: 4.14e-156 Length: 564
 Score: 169.00 Matches: 169
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 35.96% Indels: 0
 DB: Gaps: 0

US-09-993-966-7 (1-470) x BM711145 (1-564)

Qy 3 LysLeuHisSerLysProAlaAlaValCysLysArgArgLysSerProGlnGlyAspSer 22
 Db 1 AAACCTCACTCCAAAGCCGCGCGGTGTGCAACGCGAGAGAGCCGGAAGGTGACAGC 60
 Qy 23 PheAlaValSerAlaAlaTTPAlaArgLysGlyIleGlnGluTTPPileGlyArgGlnArg 42
 Db 61 TTGCGCGTGAAGCCCTGCTGCGCTCGAAGGCGATCAGAGAGTGGATCGGGAAGACAGCGC 120
 Qy 43 CysProGlyGlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGlyArgSer 62
 Db 121 TGCCTGGGCGGTGTCTCGGAGCCCGCAGACTCGGTTGGCGGAGCCATATAGCCCGAAGC 180
 Qy 63 ThrArgLysLeuValGlyAspValLeuArgAspThrLysSerGlnGlnGlnGlnGlnGln 82

Db 181 ACCGGAGCTCGTGGCGACGTGTGAGAGACAGCTCAGCGAGAGAGAGACGAC 240
Qy 83 PheArgLeuGluValAlaLeuProProGluLysThrSpGlyLeuGlySerGlyAspGlu 102
Db 241 TTTCGGCTGGAAGGGCTGCTCTCTGAGAAAGCTACGGGCTGGCAGCGGAGAGAG 300
Qy 103 LysLeuMetGluArgValSerGluProCysProGlySerLysLysGlnLeuLysPheGlu 122
Db 301 AAGAGATTGAG 360
Qy 123 GluLeuGlnCysAspValSerMetGluGluAspSerArgGlnGluTrpThrPheThrLeu 142
Db 361 GAGCTCAGTGGCAGCTGTCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Qy 143 TyrAspPheAspPheAsnGlnLysValThrArgGluAspIleThrSerLeuLeuHisThr 162
Db 421 TATGACTTTTACACAAACGCGAGAGTCAACCGAGAGAGAGAGAGAGAGAGAGAGAG 480
Qy 163 IleTyrGluValValAspSerSerVal 171
Db 481 ATCTATGAGGTGTGAGACTCTCTGTCTC 507

RESULT 15
CB215756 622 bp mRNA linear EST 06-FEB-2003
LOCUS NISC.np09e12.y1 NICHD_HS_Ut1 Homo sapiens cDNA clone IMAGE:5937502
DEFINITION 5', mRNA sequence.
ACCESSION CB215756
VERSION CB215756.1 GI:28263948
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 622)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
CDNA Library Preparation:
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
info@image.llnl.gov
Plate: LLAMA3164 row: I column: 23
Seq primer: M13RPI reverse primer (ABI).
Location/Qualifiers
1. 622
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5937502"
/sex="female"
/tissue_type="normal endometrium, late proliferative
phase, cycle day 13"
/lab_host="DH10B (T1-resistant)"
/clone_id="NICHD_HS_Ut1"
/note="Organ: uterus; Vector: pCMV-SPORT6.1.cdbd (Resgen,
Invitrogen Corporation); Site_1: NCI; Site_2: ECRV;
Cloned unidirectionally from microquantity amounts of mRNA
from normal endometrial tissue (late proliferative phase,
cycle day 13). Average insert size 1.9 kb. Library
constructed by Resgen (Invitrogen Corporation)."

ORIGIN
Alignment Scores:
Pred. No.: 3.35e-145 Length: 622
Score: 158.00 Matches: 158
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 33.62% Indels: 0
DB: 6 Gaps: 0
US-09-993-966-7 (1-470) x CB215756 (1-622)

Qy 313 ValAspProAlaSerPheHisPheLeuAspThrProIleAlaLysValSerGluLeuGln 332
Db 2 GTGACCCGGCCCTCCTTCCACTTCTTGAACCCCAATGCGCAAGCTCAGAGCTCAG 61
Qy 333 GlnArgLeuArgGlyThrGlnAspGlySerLysPheValArgSerProLysAlaGln 352
Db 62 CAACGGCTCCGGGGCACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 121
Qy 353 GlyLysSerValGlyValGlyHisValAlaArgGlyValArgAsnLysProProLeuGly 372
Db 122 GGCAGAGTGTGGGTGTGGGCGACGTGGCCAGAGGGGCAAGAAACAAGCCCTTGGGA 181
Qy 373 ProAlaIleProAlaValSerProSerAlaHisLeuAlaIleSerProAlaLeuLeuPro 392
Db 182 CCGGCATCCCTCGGTGGTCTCCCTCGGCCACCTGGCTGCCAGCCGGCCCTCTCCC 241
Qy 393 SerLeuAlaProLeuGlyHisLysLysHisLysHisIleArgAlaLysGluSerGlnGlnGly 412
Db 242 TCCCTAGCCCTCCCTCGGGCACAAGAGACACAGACCCAGACCAAGAGAGAGAGAGAG 301
Qy 413 CysArgGlyLeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlyArgGlyHisLeu 432
Db 302 TGCCTGGGCTTGCAGGACCACTGGCTTCAAGTGGCTTCTTGGGGCGGAGACACTTG 361
Qy 433 ArgGluLeuProAlaLeuValValTyrGluSerGlnAlaGlyGlnProValGlnArgHis 452
Db 362 CGGGAGCTGCCCGCTTGGTGGTGAAGAGCCAGGCCGGGAGCGGCTCCAGAGACAT 421
Qy 453 GluHisHisHisHisHisGluHisHisHisHisIleTyrHisIlePheTyrGlnThr 470
Db 422 GAGCACCACACACACATGAACATCACACCATTAACACCACTTCAACAGACA 475

Search completed: December 30, 2004, 13:42:16
Job time : 5653 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OW protein - nucleic search, using frame_plus_p2n model

Run on: December 30, 2004, 06:25:22 ; Search time 8161 Seconds
(without alignments)
2723.460 Million cell updates/sec

Title: US-09-993-966-7

Perfect score: 470

Sequence: 1 MGKLSKPAVCKRRSPG.....RHEHHHHHHHHYHFIQT 470

Scoring table:
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9045947

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O/cn2.1/USPTO.spool.p/US0993966/runat.29122004.130829.9380/app.query.fasta.1.647
-DB=GenEmbl -OPMT=fastest -SUFFIX=01g.rge -MINMATCH=0.1 -LOOPC=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=01iso -TRANS=human40.cdi -LIST=45
-DOCALL=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTPMT=ptc
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0993966 @CN2.1.1.7406 @runat.29122004.130829.9380 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEOUTERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -MARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FAGAPOP=6
-FAGEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl:.*
1: gb_ba:.*
2: gb_htg:.*
3: gb_in:.*
4: gb_om:.*
5: gb_ov:.*
6: gb_pat:.*
7: gb_ph:.*
8: gb_pl:.*
9: gb_dr:.*
10: gb_ro:.*
11: gb_sts:.*
12: gb_sy:.*
13: gb_un:.*
14: gb_vi:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	470	100.0	1438	9	AF358135 Homo sapi
2	470	100.0	1786	9	AY061883 Homo sapi
3	470	100.0	1788	9	AB062886 Homo sapi
4	470	100.0	2604	9	BC051288 Homo sapi

RESULT 1	AF358135	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	TITLE	AUTHORS	JOURNAL	FEATURES
383	81.5	1438	6	AR405917									
270	57.4	1053	6	CQ721681									
196	41.7	163319	9	HS403140									
196	41.7	169773	9	AC007608									
196	41.7	170791	9	AC007334									
10	88	18.7	178790	2	AC145238								
11	88	18.7	207486	2	AC145175								
12	58	12.3	194869	2	AC145003								
13	58	12.3	198189	2	AC145043								
14	49	10.4	1401	6	AX172351								
15	49	10.4	1416	10	AF343352								
16	49	10.4	1671	10	AF358134								
17	49	10.4	1765	10	BC034838								
18	46	9.8	1950	6	AX557538								
19	45	9.6	1731	6	AR405916								
20	45	9.6	2340	6	CQ741468								
21	44	9.4	181401	2	AC146663								
22	44	9.4	199277	2	AC145445								
23	37	7.9	208051	2	AC145254								
24	32	6.8	232509	2	AC134093								
25	32	6.8	240965	2	AC146054								
26	30	6.4	139145	2	AC146326								
27	26	5.5	237332	2	AC131846								
28	24	5.1	212531	2	AC145178								
29	22	4.7	2469	6	CQ843478								
30	22	4.7	2469	9	AK124375								
31	21	4.5	181105	5	BX664719								
32	17	3.6	1191	6	CQ715890								
33	17	3.6	1285	6	AR405918								
34	17	3.6	1307	6	AR405919								
35	17	3.6	1309	9	BC004940								
36	17	3.6	1682	9	AF358137								
37	17	3.6	1798	9	AB062887								
38	17	3.6	1885	6	AX780854								
39	17	3.6	1947	10	AF358136								
40	17	3.6	1958	9	BC012176								
41	17	3.6	1967	10	BC019952								
42	17	3.6	58359	2	AC110733								
43	17	3.6	99335	2	AC010446								
44	17	3.6	162641	2	AC091406								
45	17	3.6	187289	9	AC116351								

ALIGNMENTS

RESULT 1
AF358135 1438 bp mRNA linear PRI 29-MAY-2001
LOCUS
DEFINITION Homo sapiens naked cuticle-1 (NKD1) mRNA, complete cds.
ACCESSION AF358135
VERSION AF358135.1 GI:14211713
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1438)
Wharton,K.A., Jr., Zimmermann,G., Rousset,R. and Scott,M.P.
Vertebrate proteins related to Drosophila Naked Cuticle bind
Dishevelled and antagonize Wnt signaling
Dev. Biol. 234 (1), 93-106 (2001)
JOURNAL
MEDLINE
PUBMED
2 (bases 1 to 1438)
Wharton,K.A., Jr., Zimmermann,G. and Scott,M.P.
Direct Submision
TITLE
JOURNAL Submitted (08-MAR-2001) Pathology and Molecular Biology, University
of Texas Southwestern Medical School, 5323 Harry Hines Blvd.,
Dallas, TX 75390-9072, USA
FEATURES
source
1..1438
/organism="Homo sapiens"
/mol_type="mRNA"

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS 1
TITLE Molecular cloning, gene structure, and expression analyses of NKD1 and NKD2
JOURNAL Int. J. Oncol. 19 (5), 963-969 (2001)
MEDLINE 21490203
PubMed 11604995
REFERENCE 2 (bases 1 to 1788)
AUTHORS Katoh, M.
TITLE Direct Submission
JOURNAL Submitted (08-JUN-2001) Masaru Katoh, National Cancer Center Research Institute, Genetics and Cell Biology Section, Tsukiji 5-Chome, Chuo-Ku, Tokyo 104-0045, Japan (E-mail: mkatoh@ncc.go.jp, Tel: 81-3-3542-2511)
FEATURES
source
1. 1788
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="human fecal tissues"
1. 1788
/gene="NKD1"
6. 1418
/gene="NKD1"
/codon_start=1
/product="Dvl-binding protein NKD1"
/protein_id="BAB70500.1"
/db_xref="GI:16303260"
/translation="MGKLSKPAVCKRRESPPGDSFAVSAAMAKGIEEMIGRQPCGGVSGRQDLAIGTIGKSTRELVDVRLDTLSREEDDFLEVALPEPKTDGASDSEKKERVSEPCPSGSKQLKFEELQCDVSEEDSRQEWFTLYDIDNNKVRREDITSLHTTYEVDSSVNSPTSSKMLRYVLIVAPDSQS KRSVLVNOADLQEARPRAEIKPESDLRSWEKKORAPLRFQGD SRL EOGSCYHNCVDENIERNNHYLDLAGIENTYSGFGPSPSVAQSESLPRTSNPTFRSHPEBAIHIPIHKPKQGVDPASPHLDLPRIKVSFLQORLRGTQDSSGKH FVSPKAGCSVGVGHVBARANKPRLPAIPVSPAHILASPLALP SLAPLGHKHKHRAKSSQCCRGILQAPLASGVPVLGRHLELPALVVTESQAGQPVQ RHEHHHHHHHHHHHPYQ"
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 1788
Score: 470.00 Matches: 470
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-993-966-7 (1-470) x AB062886 (1-1788)
QY 1 MetGlyLeuLeuHisSerLysProAlaAlaValCysIleValArgArgIleSerProGluGly 20
Db 6 ATGGGAAACTTCACTCCAAAGCCGGCCGGTGTGCAAGCGCAGAGAGAGCCCGAAAGGT 65
QY 21 AspSerPheAlaValSerAlaAlaIlePAlaArgLysGlyIleGluGluTPrIleGlyArg 40
Db 66 GACAGCTTCCCGCGAGCGCTGCTGGGCTCGGAAGGCATCCGAGAGTGATCGGAGAGA 125
QY 41 GlnArgCysProGlyValValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGly 60
Db 126 CAGGCGTCCCGGCGGCTGTCTCGGAGCCCGGACAGCTGCGGTTGGCGGACCAATAGGC 185
QY 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80
Db 186 CGAAGCACCCGGGAGCTCGTGGCGAGCGTGTGAGAGCAAGCTCAACGAGGAAGAGAG 245
QY 81 AspAspPheArgGluGluValAlaLeuProProGluLysThrArgGlyLeuGlySerGly 100
Db 246 GAGCACTTTCGGCTGCGAAGTGGCCCTGCTCTGAGAAAGACTGACGGGCTGGCGAGCGGA 305

QY 101 AspGluLeuLysMetGluArgValSerGluProCysProGlySerLysGlnLeuLys 120
Db 306 GATGAGAAAGATGAGAGAGTGAAGGAACTCTGCCAAGCTCCAAAGAGCAAGCTGAAG 365
QY 121 PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGlnGluTPrPhe 140
Db 366 TTGAAGAGCTCCAGTGGCGAGTGTCCATGAGAGAGACGCGGCGAGGTGACCTTC 425
QY 141 ThrLeuThrAspPheAspAsnGlnLysValThrArgGluAspIleThrSerLeuLeu 160
Db 426 ACCCTGATACCTTTGACCAACAGGAGAGTACCCGAGAGACATCACAGCTTCTG 485
QY 161 HisThrIleTyrGluValValAspSerSerValAsnHisSerProThrSerSerLysMet 180
Db 486 CACACATCTATAGAGGTGGACTCTCTGTCAACACTCCCAACATCCAGAAATG 545
QY 181 LeuArgValLysLeuThrValAlaProAspGlySerGlnSerLysArgSerValLeuVal 200
Db 546 CTCGGGTAAAGCTCACCGTGGCCCGGATGGCAGGCAGCAAGAGAGCGTCTTGTC 605
QY 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGluThrLysProThrGluAspLeu 220
Db 606 AATCAGGCTGACCTGCAGAGCGCAAGGCCCGGAGCAGAGACCAAGCCACTGAGAGACTG 665
QY 221 ArgSerTrpGluLysLysGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
Db 666 CGAGAGCTGGAGAAAGAGAGCAGCGAGCCCGCTCAGGTTCAGAGGTGACACCCGCTGGAG 725
QY 241 GlnSerGlyCysTyrHisHisCysValAspGluAsnIleGluArgArgAsnHisTyrLeu 260
Db 726 CAGTGTGGCTGTACCAACCATTCGTAGTGAAGAACATCAAGAGAGAAACCACTACTTA 785
QY 261 AspLeuAlaGlyIleGluAsnTyrThrSerGlnPheGlyProGlySerProSerValAla 280
Db 786 GATCTGCCCGGATAGAAATCTACAGTCCCAATTGGAGCTGCTCCCTTCGTGGGCC 845
QY 281 GlnLysSerGluLeuProProArgThrSerAspProThrArgSerArgSerHisGluPro 300
Db 846 CAGAGGTCAAACTGGCCCCCGCCACCTCCAAITCCCATCTGATCTGCTCCATAGACCG 905
QY 301 GlnAlaIleHisIleProHisArgLysProGlnGlyValAspProAlaSerPheHisPhe 320
Db 906 GAAGCCATCCACATCCACACCAAGGCCCAAGGCGGTGACCCGCGCTCTCCACTTC 965
QY 321 LeuAspThrProIleAlaLysValSerGluLeuGlnGlnArgLeuArgGlyThrGlnAsp 340
Db 966 CTTGACACCCCAATCGCCAAAGTCTTCAGAGCTCCAGCAAGGCTCCGGGACCCAGAGAC 1025
QY 341 GlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValGlyHis 360
Db 1026 GGGAGCAAGCACTTGTGAGGTCCCCCAAGGCCCAAGGAGTGGGTGGGCCAC 1085
QY 361 ValAlaArgGlyValaArgAsnLysProProLeuGlyProAlaIleProAlaValSerPro 380
Db 1086 GTGGCCAGAGGGCCAGAAACAAGCCCCCTGTGGAGCCCCCATCTCTGGGTGTCCCCC 1145
QY 381 SerAlaHisLeuAlaIleAspProAlaLeuLeuProSerSerLeuAlaProLeuGlyHisLys 400
Db 1146 TCGGCCCACTGGCTGCAAGCCCGGCTCTCTCCCTCTAGCCCCCTCGGGCAAG 1205
QY 401 LysHisLysHisArgAlaLysGlnSerGlnGlnGlyCysArgGlyLeuGlnAlaProLeu 420
Db 1206 AAGCACAAGCACCGAGCCAGAGAGAGCCAGAGGCTGCCGGGCTCGAGGACCAACTG 1265
QY 421 AlaSerGlyGlyProValLeuGlyArgGlnHisLeuArgGluLeuProAlaLeuValVal 440
Db 1266 GCTCAGAGTGGCCCTGTCTGTGGGCGGAGCACTGCGGAGGTGCGCCGCTTGTTGTG 1325
QY 441 TyrGluSerGlnAlaGlyGlnProValGlnArgHisGluHisHisHisHisGluHis 460
Db 1326 TATGAGAGCAGGCGGGCAGCGGTCCAGAGACATGAGCAACCAACCAATGAACAT 1385
QY 461 HisHisIleTyrHisIlePheTyrGlnThr 470

Db 1386 CACCACATTACCACTCTTACCAACA 1415

RESULT 4
BC051288
LOCUS
DEFINITION Homo sapiens naked cuticle homolog 1 (Drosophila), mRNA (cDNA clone MGC:59666 IMAGE:6646089), complete cds.

ACCESSION BC051288
VERSION
KEYWORDS
SOURCE MGC.
ORGANISM Homo sapiens (human)

REFERENCE 1 (bases 1 to 2604)
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heien, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hult, S.W., Villalón, D.R., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marz, M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL 12477932
PUBMED 2 (bases 1 to 2604)
REFERENCE Strausberg, R.
AUTHORS Direct Submission
TITLE Submitted (25-APR-2003) National Institutes of Health, Mammalian Gene Collection (MGC) Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland:
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mgc@nri.nih.gov
Akhtar, N., Ayala, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakeley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Hashiguchi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaapi, R., Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCluskey, J.C., McDowell, J., Pearson, R., Stancijop, S., Thomas, P.J., Touchman, J.W., Tsueng, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

FEATURES
source
1..2604
/organism="Homo sapiens"

gene
CDS
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:59666 IMAGE:6646089"
/issue="Ovary, teratocarcinoma"
/issue_type="Ovary, teratocarcinoma"
/clone_id="NIH MGC_109"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
1..2604
/gene="NKD1"
/db_xref="LOCUSID:85407"
/db_xref="MIM:607851"
225..1637
/gene="NKD1"
/codon_start=1
/product="naked cuticle homolog 1"
/protein_id="AAH51288.1"
/db_xref="GI:30410966"
/db_xref="LOCUSID:85407"
/db_xref="MIM:607851"
/translation="MGKLSKPAVCKRESPEGDSFVSAAMARKIEWIGRQCP
GVSGRPQLRLAGTIGKSTRELVDVLRDTLSSEEDPFLVLAPEKTDGLSSGB
KMERVSEPCPSKQKLEBELQDVMSBEDPFWTFLYDPDNGKVRBEDITSL
HTIYEVDSVNSPTSSKMLRYVLAIVAPGSGSKSVLVNQADLQARPAETKPE
DLRSEKKORAPLRFQDSRLLEQSGCYHCVDENIBERNHYLDLAGIENTSOPGDS
PSVAQSELPSPRNSRSHPEALHITPHRQCVDPASFRHLDPIAKVSLQOR
LRGTQDSKHFVSPKQGSVGVGVARARNPPLGPAIPVPSAHILASFLAP
SLAPLHKRHKRHKRSQQRGLQAPLASGVPVLRHLRELPAVVESQAGQPVQ
RHHHHHHHHHHHHHYQY"

ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 2604
Score: 470.00 Matches: 470
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-09-993-966-7 (1-470) x BC051288 (1-2604)

QY 1 MercGlyLysLeuHisSerLysProAlaAlaValCybYsArgArgGluSerProGluGly 20
Db 225 ATGGGAAATCTTCACTCCAGCCGCCCGCTGTGCAAGCGAGAGAGAGAGAGAGT 284
QY 21 AspSerPheAlaValSerAlaAlaATPAlaArgLysGlyLleGluGluTPTPLeGlyArg 40
Db 285 GACGAGCTTCGCCGTGACCGCTGCTGGGCTCGGAGAGGAGATCGAGAGTGGATGGGAGA 344
QY 41 GlnArgCybProGlyGlyValSerGlyProArgGluLeuArgLeuAlaGlyThrLleGly 60
Db 345 CAGGCTGCGCCGGCGGTGTCTCGGAGCCCGACAGCTGCGTGGCGGACACCATAGGC 404
QY 61 ArgSerThrArgGluLeuValAlaLeuProProGluLysThrAspGlyLeuGlySerGly 100
Db 405 CGAAGCACCCGGAGCTCGTGGCGAGCTTTAGAGACACGCTACCGAGAGAGAGAG 464
QY 81 AspAspPheArgLeuGluValAlaLeuProProGluLysThrAspGlyLeuGlySerGly 100
Db 465 GAGGACTTTCGCTGGAAGAGCGCCCTGCTCGAGAGAGCTGAGCGGCTGGGACGCGA 524
QY 101 ArgGluLysLeuMetGluValSerGluProCybProGlySerLysLysGluLeuLys 120
Db 525 GATGAGAGAGAGTGGAG 584
QY 121 PheGluGluLeuGlnCybAspValSerMetGluGluAspSerArgGlnGluTPTPThrPhe 140
Db 585 TTGGAAGAGCTCCAGTCCGACGCTGCTCATGAGAGAGACAGCCGCGAGAGTGCACCTTC 644
QY 141 ThrLeuTyrrAspPheAspAsnAnglyLysValThrArgGluLysPierSerLeuLeu 160
Db 645 ACCCTGATGACTTTGACACCAACGCGAGAGTACCCGAGAGAGACATCACGAGCTTCTG 704
QY 161 HisThrIleTyrrGluValAlaAspSerSerValAsnHisSerProThrSerSerLysMet 180

Db 705 CACGCCATCTATGAGGTGGTGAATCTCTGTGTCAACCACTCCCAACATCCAGCAATG 764
Qy 181 LeuArgValIysLeuThrValAlaProAspGlySerGlnSerIlyAspSerValLeuVal 200
Db 765 CTGGGGGTAAAGGCTCACCGTGGCCCCGATGGCCAGCAAGCAAGAGAGAGGCTCTTGTC 824
Qy 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGlnThrIlyProThrGluAspLeu 220
Db 825 AATCAGGCTAATCTGCAAGCGCAAGGCCCGGAGCCAGACCAAGCCCACTGAGAGCTTG 884
Qy 221 ArgSerTrpGluIlySerGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
Db 885 CGAGGCTGGAGAAAGAGAGAGAGAGCCCGCTGATTCAGAGGTGACAGCCGCTGGAG 944
Qy 241 GlnSerGlyCysTrpThrShiSerCysValAspGluAsnIleGluIleArgAlaGlnHis 260
Db 945 CAGTCTGGCTGCTACCACTTCGCTGATGAGAAACAATCGAGAGAGAAACCACTACTTA 1004
Qy 261 AspLeuAlaGlyIleGluAsnTrpThrSerGlnPheGlyProGlySerProSerValAla 280
Db 1005 GATCTCCCGGAGATAGAAACCTACAGTCCCAATTTGGGCTGGCTCCCTTCCTGGGCC 1064
Qy 281 GlnIlySerGluLeuProArgTrpThrSerAsnProThrArgSerArgSerHisGluPro 300
Db 1065 CAGAGCTCAGAGACTGCCCGCCCGCACTCCATCCATCCATGCTGCTCCCATGAGCCG 1124
Qy 301 GlnAlaIleHisIleProHisArgIlyProGlnIlyValAspProAlaSerPheHisPhe 320
Db 1125 GAAGCCATCCACATCCACACCGAAAGCCCAAGCCGTGAGCCCGGCTCCCTTCACCTTC 1184
Qy 321 LeuAspThrProIleAlaIlyValSerGluLeuGlnIleArgLeuArgIlyThrGluAsp 340
Db 1185 CTGTGACACCCCAATCCGCAAGGTTCTCAAGCTTCAAGACCGACTCCGGGAGCAACAGAC 1244
Qy 341 GlySerIlyShiSphenValArgSerProIlySerValGlnIlySerValGlyValGlyHis 360
Db 1245 GGGAGCAAGCACTTGTGAGGTGCCCAAGGCCCAAGGCAAGAGTGTGGGTGGGCGAC 1304
Qy 361 ValAlaArgGlyAlaArgAsnIlyProProLeuGlyProAlaIleProAlaValSerPro 380
Db 1305 GTGGCCAGAGGGGCAAAACCAAGCCCCCTGTGGAGCCCGCCATCCCTGGGTGTCCCC 1364
Qy 381 SerAlaHisLeuAlaAspSerProAlaLeuLeuProSerLeuAlaProLeuGlyHis 400
Db 1365 TCCGCCCACTGGTGTGCAAGCCCGGCTCTCTCCCTCCCTAGCCCCCTCGGGCAAG 1424
Qy 401 LysHisIlyShiSargAlaIlySerGlnGlnIlyCysArgGlyLeuGlnAlaProLeu 420
Db 1425 AAGCAACAAGCACCGAGCCAGAGAGCCAGAGGCTGCCGGGCTCGAGGCAACAATG 1484
Qy 421 AlaSerGlyIlyProValLeuGlyArgGluHisIleLeuArgGlyLeuProAlaLeuVal 440
Db 1485 GCGCTCAGGTGGCTGTCTGTGGGCGGAGAGCACTGGGGAGTGGCCGCTTGTGTG 1544
Qy 441 TrpGluSerGlnAlaGlyIlyProValGlnArgHisGlnHisIlyShiShiShiGlnHis 460
Db 1545 TATAGAGCCAGGCGGAGAGCCGCTCCAGACATGAGACCAACCACTATGAACAT 1604
Qy 461 HisHisIlyTrpHisIlySphenTrpGlnThr 470
Db 1605 CACCACCATTAACCACTTCAACAGACA 1634

RESULT 5
AR405917 1438 bp DNA linear PAT 18-DEC-2003
LOCUS DEFINITION Sequence 5 from patent US 6630323.
ACCESSION AR405917
VERSION AR405917.1 GI:40154965
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1438)
AUTHORS Scott,M., Zeng,M. and Wharton,K.
TITLE Naked cuticle genes and their uses
JOURNAL Patent: US 6630323-A 5 07-OCT-2003;
FEATURES
source Location/Qualifiers
1..1438
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 1438
Score: 383.00 Matches: 469
Percent Similarity: 99.58% Conservative: 0
Best Local Similarity: 99.58% Mismatches: 1
Query Match: 81.49% Indels: 2
DB: Gaps: 0
US-09-993-966-7 (1-470) x AR405917 (1-1438)
Qy 1 MetGlyIlyLeuHisSerIlySerProAlaAlaValCysIlySargArgGluSerProGluGly 20
Db 8 ATGGGAAACTTCACTTCAAGCCGCGCGGTGGCAAGCGAGAGAGCCGGAAGGT 67
Qy 21 AspSerPheAlaValSerAlaAlaTrpAlaArgIlyGlyIleGluGluTrpIleGlyArg 40
Db 68 GACAGCTTCCCGCTGAGCGCTGCTGGGCTCGGAAGGCATTCGAGAGTGGATCGGAGA 127
Qy 41 GlnArgCysProGlyIlyValSerGlyProArgGlnLeuArgLeuAlaGlyThrIleGly 60
Db 128 CAGGCTGCCCGGCGGTGTCTGGGACCCCGACAGCTGGGTTGGCGGACCATTAGGC 187
Qy 61 ArgSerThrArgGluLeuValGlyAspValLeuArgAspThrLeuSerGluGluGlu 80
Db 188 CGAAGCACCCGGAGGCTGTGGGCGAGGTGTGAGACACGCTCAGCCGAGAAAGAGAG 247
Qy 81 AspAspPheArgLeuGluValAlaLeuProProGluIlySerThrAspGlyLeuGlySerGly 100
Db 248 GACGACTTTCGGCTGGAAAGTGGCCCTGCTCTGAGAAAGCTGACGGGCTGGGAGCGGA 307
Qy 101 AspGluIlySphenMetGluArgValSerGluProCysProGlySerIlySphenLeu 120
Db 308 GATGAGAAAGATGAGAGAGAGAGAACCTTGCCAGGCTTCCAGAAAGCAAGCTTGAAG 367
Qy 121 PheGluGluLeuGlnCysAspValSerMetGluGluAspSerArgGlnIlyTrpThrPhe 140
Db 368 TTGGAAGCTCAAGTGCAGCTGTCCATGAGAGAGACCGGAGAGTGAACCTTC 427
Qy 141 ThrLeuTrpAspPheAspAsnAngIlyValThrArgGluAspIleThrSerLeuLeu 160
Db 428 ACCCTGATATGACTTTGACAAACAGCAAGTCAACCGAGAGCATCACAGCTTCTG 487
Qy 161 HisThrIleTrpGluValValAspSerSerValAsnHisSerProThrSerSerIlyMet 180
Db 488 CACACCATATAGAGGTGTGACTCTCTGTCAACCACTCCCAACATCAGCAAGATG 547
Qy 181 LeuArgValIlyLeuThrValAlaProAspGlySerGlnSerIlySargSerValLeuVal 200
Db 548 CTGGGGTAAAGCTCACCGTGGCCCCCGATGGAGCGAGAGCAAGAGAGAGCGTCTTGTC 607
Qy 201 AsnGlnAlaAspLeuGlnSerAlaArgProArgAlaGlnThrIlyProThrGluAspLeu 220
Db 608 AATCAGGCTAATCTGCAAGCGCAAGGCCCGGAGCCAGACCAAGCCCACTGAGAGCTG 667
Qy 221 ArgSerTrpGluIlySerGlnArgAlaProLeuArgPheGlnGlyAspSerArgLeuGlu 240
Db 668 CGAGGCTGGAGAAAGAGAGAGAGCCCGCTGATTCAGAGGTGACAGCCGCTGGAG 727
Qy 241 GlnSerGlyCysTrpThrShiSerCysValAspGluAsnIleGluIleArgAlaGlnHis 260
Db 728 CAGTCTGGCTGCTACCACTTCGCTGATGAGAAACAATCGAGAGAGAAACCACTACTTA 787
Qy 261 AspLeuAlaGlyIleGluAsnTrpThrSerGlnPheGlyProGlySerProSerValAla 280

```

Db      788  GATCTGCGCGGGATGAAAATCAAGTCCCAATTGGGCTGCTCCCTCCGTTGACC 847
Qy      281  GlnLysSerGluLeuProArgThrSer-AsnProThrArgSerArgSerHisGluPro 300
Db      848  CAGAAAGTCAGAAATGCCCCCCCCACCTT-CAATCCCACTGATCTGCTCCCATGAGCC 906
Qy      300  oGluAlaIleHisIleProHisArgGlySerProGlnGlyValAspProAlaSerPheHisPhe 320
Db      907  GGAAGCCATCCACATCCACACCGAAAGCCCAAGGGGTGAGCCCGGCTCTCCCTCC 966
Qy      320  eLeuArgThrProGlnAlaAlaValSerGluLeuGlnGlnArgLeuArgGlyThrGlnAs 340
Db      967  CTTTGACACCCCAATGCGCAAGGTCTAGAGCTCCAGCAACGGCTCCGGGGACTCAGGA 1026
Qy      340  pGlySerLysHisPheValArgSerProLysAlaGlnGlyLysSerValGlyValGlyHis 360
Db      1027  CGGAGCAAGCACTTGTGTGAGTCCCCCAAGGCCCAAGGCAAGAGTGTGGGTGTGGGCA 1086
Qy      360  eValAlaArgGlyValArgHisLysProProLeuGlyProAlaIleProAlaValSerPr 380
Db      1087  CGTGCCAGAGGGGCAAGAAAGCCCTCTGGGACCCGCCATCCCTGCGGTGCC 1146
Qy      380  oSerAlaHisLeuAlaAlaSerProAlaLeuLeuProSerLeuAlaProLeuGlyHisLys 400
Db      1147  CTCGCCCACTGCTGCTGCGCAAGCCCGGCTCTCTCCCTCCCTAGCCCCCTCCGGGCAAA 1206
Qy      400  sLysHisLysHisArgAlaLysGluSerGlnGlnGlyCysArgGlyLeuGlnAlaProLe 420
Db      1207  GAAGCAAGACCAAGGCAAGGAGGCAAGCGGCTGCGGGGCTGCAAGGCAACCACT 1266
Qy      420  uAlaSerGlyGlyProValLeuGlyArgGlyHisLysArgGlyLeuProAlaLeuValVa 440
Db      1267  GGGCTAGGTGGCCCTGCTCTGGGGGGGAGCACTGCGGAGCTGCCCCCTTGGGTGT 1326
Qy      440  lTyrgLysSerGlnAlaGlyGlnProValGlnArgHisGluHisHisHisHisGluHis 460
Db      1327  GRTTGAGAGCCAGGCGCGGCGAGCCGCTCCAGAGACATGAGCACCAACCAATGAA 1386
Qy      460  sHisHisHisLysTyrgHisHisPheTyrgLysHis 470
Db      1387  TCACCACCATTTACCAACCACTTACAGAGACA 1417

RESULT 6
CQ721681  CQ721681  1053 bp  DNA  linear  PAT 03-FEB-2004
LOCUS      Sequence 7615 from Patent WO2068579.
DEFINITION CQ721681
ACCESSION  CQ721681
VERSION     CQ721681.1 GI:42282538
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
  TITLE     Kites, such as nucleic acid arrays, comprising a majority of
            humenexons or transcripts, for detecting expression and other uses
            thereof
  JOURNAL   Patent: WO 02068579-A 7615 06-SEP-2002;
            PE Corporation (NY) (US)
FEATURES
  source     Location/Qualifiers
            1..1053
               /organism="Homo sapiens"
               /mol_type="unassigned DNA"
               /db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.:      2,366-264      Length:      1053
Score:          270.00         Matches:      270
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0

```

```

Query Match: 57.45%      Indels: 0
DB:          6          Gaps: 0
US-09-993-966-7 (1-470) x CQ721681 (1-1053)

Qy      87  ValAlaLeuProProGluLysThrArgGlyLeuGlySerGlyValAspGluLysMetGlu 106
Db      82  GTGGCCCTGCTCTTGAGAGACTGAGGGCTGGGAGCGGAGTGAAGAGAGATGAG 141
Qy      107  ArgValSerGluProCysProGlySerLysGlnLeuLysPheGlnGluLeuGlnCys 126
Db      142  AGAGTAGGCAACCTGCGCCAGGCTCCAGAGACAGTGAAGTTTGAAGGCTCCAGTGC 201
Qy      127  AspValSerMetGluLysAspSerArgGlnGlnTTrThrPheThrLeuTyraAspPheAsp 146
Db      202  GACGTGTCCATGAGAGGAGGACAGCCGCGAGAGATGACCTTCAACCTGTGTGACTTTGAC 261
Qy      147  AsnAsnGlyLysValThrArgGluAspLysThrSerLeuLeuHisThrLysTyrgLysVal 166
Db      262  AACACGGCAAGGTCAACCCAGAGAGACATACAGCTTGTCTGCAACCATCTTAGAGTGT 321
Qy      167  ValAspSerSerValAsnHisSerProThrSerSerLysMetLeuArgValLysLeuThr 186
Db      322  GTGGACTCCTCTGTCAACCACTCCCAACATCCAGAGAGATGCTGGCGGTAAAGCTCAC 381
Qy      187  ValAlaProAspGlySerGlnSerLysArgSerValLeuValAsnGlnAlaAspLeuGln 206
Db      382  GTGGCCCCCATGTGGCAGCCAGCAAGAGAGAGGTCTCTGTCAATCAGGCTGACCTGCAG 441
Qy      207  SerAlaArgProArgAlaGluThrLysProThrGluAspLeuArgSerTrpGluLysVal 226
Db      442  AGGCAAGGCCCCGAGAGAGAGACCAAGCCCATGAGAGCTTGCGAGCTGGAGAGAGAG 501
Qy      227  GlnArgAlaProLeuArgPheGlnGlnLysAspSerArgLeuGlnLysSerGlyCysTyrgHis 246
Db      502  CACGAGAGCCCGCTCAGGTTCCAGGGTGACAGCGGCTGGAGCAGTCTGCTCTACAC 561
Qy      247  HisCysValAspGluAsnIleGluArgArgHisLysTyrgLeuAspLeuAlaGlyLysGlu 266
Db      562  CATTTGGTGAATGAGAAACATCGAGAGAGAAACACTACTTAGATCTCGCGGGATGAA 621
Qy      267  AsnTyrgThrSerGlnPheGlyProGlySerProSerValAlaGlnLysSerGluLeuPro 286
Db      622  AACTAACAGTCCCAATTGGGCTGAGCTCCCTTCCTGAGCCAGAGATGAGAACTGCCC 681
Qy      287  ProArgThrSerAsnProThrArgSerArgSerHisGluProGluAlaIleHisLysPro 306
Db      682  CCGGCACTCCCAATCCCACTGATCTCGCTCCAGAGAGCCGAGAGCCATCCATCCCA 741
Qy      307  HisArgLysProGlnGlyValAspProAlaSerPheHisPheLeuAspThrProIleAla 326
Db      742  CACCGAAGGCCCAAGCGGTGAGCCCGGCTCTCTCCACTTCCATTCATCCCAATCCGCC 801
Qy      327  LysValSerGluLeuGlnGlnArgLeuArgGlyThrGlnAspGlySerLysHisPheVal 346
Db      802  AAGGTCTCAGAGTCTCAGAGACGGCTCCGGGGACCCAGGACGGAGCAAGCACTTGTG 861
Qy      347  ArgSerProLysAlaGlnGlyLysSerVal 356
Db      862  AGGTCCCGCAAGGCCCAAGGCAAGAGTGTG 891

RESULT 7
HSAJ03140  HSAJ03140  16339 bp  DNA  linear  PRI 03-APR-2004
LOCUS      Homo sapiens NOD2 gene for LRR-containing protein, exons 1-11.
DEFINITION HSAJ03140
ACCESSION  HSAJ03140
VERSION     AJ303140.1 GI:14488148
KEYWORDS    LRR-containing protein; NOD2 gene.
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1

```


AUTHORS Hugot, J. P., Chamailhard, M., Zouali, H., Lesage, S., Cezard, J. P., Belatche, J., Almer, S., Tyek, C., O'Morain, C. A., Gassull, M., Binder, V., Finkel, Y., Cortot, A., Modigliani, R., Laurent-Puig, P., Gower-Rousseau, C., Macry, J., Colomdel, J. F., Sabatou, M. and Thomas, G.

TITLE Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease

JOURNAL Nature 411 (6837), 599-603 (2001)

MEDLINE 21279172

PUBMED 11385576

REFERENCE 2 (bases 1 to 163319)

AUTHORS Zouali, H.

TITLE Direct Submission

JOURNAL Submitted (02-JAN-2001) Zouali, H., Genetics, Fondation Jean Dausset-CEPH, 27 Rue Juliette Dodu, 75010 Paris France, FRANCE

FEATURES

source

1. 163319

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="16"

/map="16q12-13"

/clone_1lb="hd87b10"

/clone_1lb="CEPH Human BAC library"

98302. 131994

/gene="NOD2"

join(98302. 98906,106807. 106912,109510. 111325, 115517. 115600,115821. 115904,118855. 118938, 121552. 121635,122230. 122313,124418. 124501, 128745. 128828,130674. 131994)

/gene="NOD2"

98302. 98906

/gene="NOD2"

/number=1

join(98448. 98906,106807. 106912,109510. 111325, 115517. 115600,115821. 115904,118855. 118938, 121552. 121635,122230. 122313,124418. 124501, 128745. 128828,130674. 130746)

/gene="NOD2"

/function="putative NF-kB activator"

/note="protein of the CARD family"

/codon_start=1

/product="LRR-containing protein"

/protein_id="CAC42117.1"

/db_xref="GI:14488149"

/db_xref="GOA:Q9HC29"

/db_xref="Sw188-ProC:Q9HC29"

/translation="MCSQEAFOAQRSQLVELLVSGSLGEFESVLDMLSWELSWEDY EGFHLGQPLSHLARLDITVMKGTWACQKLI AAQEAQDSQSPKLGCMPSHSH PARDIOSHRPAIVRSLRHSHVEMLDLAWERGSPVSOYCDRIPLPIPPSGRARLDL ATYKANGLAFLIOHVELPPLALPIBAATCKTKAKTKTTVSAGSRFLSTYDGAET LCLEDITYTENNVLEWADVAGSPQKSPATLGLELFTSPGHLNDADTVLVVEGAS GKSTLLQRLHLMAAGQDFQELFVFPFSROLDQCAKPLSVRTLFEHCCMPVQGE DIQDLIDHPDRVLTFTGDFDEFEKFTDEBERHSPPDPSPVQLFENLQGNILKNA RKVTSRPAVSAFLRKYIKTEFNLKFSQOGETLYRKHNHSGVADRLIRLQETS ALHGTCILPVFSMWKSCHOELLQBGSGSKTTPDWYLLIHFLLHATPRDSASQTL GPSTLGRRLPTLHLKRLALMGCMCTTYSAAQDLQAQVSPDDISGFIYRAKVVP GSTAPLEFLHTTQCFFPAFYLAISADVPPALRLHLFNCGRPGNSPPARLLPTVCIQA SBGSDSVAALLQAEPHNIQITAAFLGILSRHWGLAEQCTSEKALRRQACAM CLARSLKHFHSIPPAFGBAKSVAMPGFIMLIRSLYEQEELAKARGLNVGHL KLTFCGVPTECCALAFVLOHLRRPVALODISVGLDIPCLGVCVKAYLIRN NISDRGICKLIEGALHCEOLKTLFNNKLTIDCGAHSMATLACRONFLARLGNVYI TAAQAQVLAAGLRGNTSLQFLGFWGNVGBGQAQDLAALQDHOSLWLSLVGNNTIS VQAQALMLAKNVMELECLIEENHLODBEVCSLAEGLKKNSSIKTLKLSNNCTTYIG AEAALQALERNDDITLVEWLRGNTFSLBEVDKGRDRTLLL"

98907. 106806

/gene="NOD2"

/number=1

106807. 106912

/gene="NOD2"

/number=2

106913. 109509

/gene="NOD2"

/number=2

109510. 111325

/gene="NOD2"

/number=3

111326. 115516

/gene="NOD2"

/number=3

115517. 115600

/gene="NOD2"

/number=4

115601. 115820

/gene="NOD2"

/number=4

115821. 115904

/gene="NOD2"

/number=5

115905. 118854

/gene="NOD2"

/number=5

118855. 118938

/gene="NOD2"

/number=6

118939. 121551

/gene="NOD2"

/number=6

121552. 121635

/gene="NOD2"

/number=7

121636. 122229

/gene="NOD2"

/number=7

122230. 122313

/gene="NOD2"

/number=8

122314. 124417

/gene="NOD2"

/number=8

124418. 124501

/gene="NOD2"

/number=9

124502. 128744

/gene="NOD2"

/number=9

128745. 128828

/gene="NOD2"

/number=10

128829. 130673

/gene="NOD2"

/number=10

130674. 131994

/gene="NOD2"

/number=11

ORIGIN

Alignment Scores:

Pred. No.: 4, 41e-187

Score: 196.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 41.70%

DB: 9

US-09-993-966-7 (1-470) x HSA303140 (1-163319)

Qy 275 GlysSerProSerValAlaGlnIySserGluLeuProProArgThrSerAspProThrArg 294

Db 31966 GGCTCCCTCCCGGGCCGAGAGTCAAGACTGCCCGCACCTCCCAATCCCACTCGA 32025

Qy 295 SerArgSerHisGluProGluAlaIleHisIleProHisArgGlySerProGlnGlyValAsp 314

Db 32026 TCTCGCTCCCATGAGCCGAGGACCATCCACATCCACACCGAAGAGCCGAGGCTGAC 32085

Qy 315 ProLaserPheHisPheLeuAspThrProIleAlaIyValSerGluLeuGlnGlnArg 334


```

Db      32086 CCGGGCTCTTCACATTCCTTGACACCCCAATGCCAAGCTCTAGAGCTCCAGACCGG 32145
Qy      335 LeuArgGlyThrGlnaPrgGlySerLyshPheValaArgSerProLyalaGlnGlyLys 354
Db      32146 CTCGGGGGACCCAGGACGGGAGCAACATTTGTGAGGTCCCCCAAGGCCCGAGGGGAG 32205
Qy      355 SerValGlyValaGlyshValaAlaArgGlyAlaArgGlnLyPProPoleuGlyProAla 374
Db      32206 AGGTGGGTGTGGGCCACGTGGCCAGGGGCAAGAAACAGCCCTCTGGGACCCGGCC 32265
Qy      375 IleProAlaValaSerProSerAlaHshIleuAlaAlaSerProAlaLeuLeuProSerLeu 394
Db      32266 ATCCCTGGGATGTCCCTCCGCCACCTGAGCTGCCAGCCCGGCTCTCTCCCTCCCTA 32325
Qy      395 AlaProLeuGlyshLyshLyshLyshLyshLyshLyshLyshLyshLyshLyshLyshLysh 414
Db      32326 GCGCCCTCGGGGACCAAGAACCAAGCAACGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGG 32385
Qy      415 GlyLeuGlnAlaProLeuAlaSerGlyGlyProValaLeuGlyArgGlyLysIleuArgGly 434
Db      32386 GGGCTGAGGACACCACTGGCTCAGGTGGCTGTCTCTGGGGGCGGAGCACTGGCGGAG 32445
Qy      435 LeuProAlaLeuValaValaTyrgLysSerGlnAlaGlyGlnProValaGlnArgHshGlnHsh 454
Db      32446 CTCGCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 32505
Qy      455 HshHshHshHshHshHshHshHshHshHshHshHshHshHshHshHshHshHshHshHshHsh 470
Db      32506 CACACACACATGACATCACTACCACTTACCACTTACCACTTACCACTTACCACTTACCACT 32553

RESULT 8
AC007608/c 169773 bp DNA linear PRI 19-MAR-2003
LOCUS Homo sapiens chromosome 16 clone RP11-401P9, complete sequence.
DEFINITION AC007608
AC007608.7 GI:29124038
KEYWORDS HTG.
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 169773)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
Direct Submission
Unpublished
2 (bases 1 to 169773)
Bruce, D., Mundt, M., Deggett, N., Munk, C., Saunders, E., Robinson, D.,
Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,
Bryant, J., Tesmer, J., Meinke, L., Longmire, J., White, S., Tatum, O.,
Campbell, C., Fawcett, J., Maltbie, M., Buesed, M., Sutherland, R.,
McMurry, K., Han, C. and Deaven, L.
Direct Submission
Submitted (20-MAY-1999) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
3 (bases 1 to 169773)
DOE Joint Genome Institute.
Direct Submission
Submitted (04-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 169773)
DOE Joint Genome Institute.
Direct Submission
Submitted (11-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
5 (bases 1 to 169773)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
Direct Submission
Submitted (19-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA

```

```

COMMENT On Mar 19, 2003 this sequence version replaced gi:18129388.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www-ehg.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated total Number of Errors is 0.
FEATURES
source
1. 169773
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-401P9"
ORIGIN
Alignment Scores:
Pred. No.: 4.57e-187 Length: 169773
Score: 196.00 Matches: 196
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 41.70% Indels: 0
DB: 9 Gaps: 0
US-09-993-966-7 (1-470) x AC007608 (1-169773)
Qy      275 GlySerProSerValaGlnLyshSerGlyLeuProProArgThrSerAenProThrArg 294
Db      78524 GGGCTCCCTTCGATGGGCGCAAGATGACAGACTGCCCGGACCTCCATCCACACTCGA 78465
Qy      295 SerArgSerHshGlnProGlnAlaIleHshIleProHshArgLyshProGlnGlyValaAsp 314
Db      78464 TCTCGCTCCCATGAGCGGAGGACATCCATCCACACCGAAAGCCCAAGGGGTGGAC 78405
Qy      315 ProAlaSerPheHshPheLeuAepThrProAlaIleValaSerGlyLeuGlnGlnArg 334
Db      78404 CCGGCTCTCTTCACATCTTGTGACACCCCAATGCGCAAGGTCTTAGAGCTCCAGCAACGG 78345
Qy      335 LeuArgGlyThrGlnaPrgGlySerLyshPheValaArgSerProLyalaGlnGlyLys 354
Db      78344 CTCGGGGGACCCAGGACGGGAGCAACATTTGTGAGGTCCCCCAAGGCCCGAGGGGAG 78285
Qy      355 SerValGlyValaGlyshValaAlaArgGlyAlaArgGlnLyPProPoleuGlyProAla 374
Db      78284 AGTGGGTGTGGGCCACGTGGCCAGGGGCAAGAAACAGCCCTCTGGGACCCGGCC 78225
Qy      375 IleProAlaValaSerProSerAlaHshIleuAlaAlaSerProAlaLeuLeuProSerLeu 394
Db      78224 ATCCCTGGGATGTCCCTCCGCCACCTGAGCTGCCAGCCCGGCTCTCTCCCTCCCTA 78165
Qy      395 AlaProLeuGlyshLyshLyshLyshLyshLyshLyshLyshLyshLyshLyshLyshLysh 414
Db      78164 GCGCCCTCGGGGACCAAGAACCAAGCAACGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGG 78105
Qy      415 GlyLeuGlnAlaProLeuAlaSerGlyGlyProValaLeuGlyArgGlyLysIleuArgGly 434
Db      78104 GGGCTGAGGACACCACTGGCTCAGGTGGCTGTCTCTGGGGGCGGAGCACTGGCGGAG 78045
Qy      435 LeuProAlaLeuValaValaTyrgLysSerGlnAlaGlyGlnProValaGlnArgHshGlnHsh 454
Db      78044 CTCGCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 77985
Qy      455 HshHshHshHshHshHshHshHshHshHshHshHshHshHshHshHshHshHshHshHshHsh 470
Db      77984 CACACACACATGACATCACTACCACTTACCACTTACCACTTACCACTTACCACTTACCA 77927

RESULT 9
AC007334/c 170791 bp DNA linear PRI 13-MAR-2003
LOCUS Homo sapiens chromosome 16 clone RP11-147B17, complete sequence.
DEFINITION AC007334
AC007334.8 GI:28933524
VERSION

```

KEYWORDS	HTG.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
REFERENCE	1 (bases 1 to 170791)
TITLE	DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.
JOURNAL	Direct Submission
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 170791)
AUTHORS	Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L., Bryant,U., Tesner,J., Meinicke,L., Longmire,U., White,S., Tatum,O., Campbell,C., Pawcett,J., Maltbe,M., Bussod,M., Sutherland,R., McMurry,K., Han,C. and Deaven,L.
JOURNAL	Direct Submission
REFERENCE	Submitted (21-APR-1999) Center for Human Genome Studies, DOE Joint Genome Institute, Los Alamos National Laboratory, MS M88e, Los Alamos, NM 87545, USA
AUTHORS	3 (bases 1 to 170791)
TITLE	DOE Joint Genome Institute.
JOURNAL	Direct Submission
REFERENCE	Submitted (24-AUG-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94599, USA
AUTHORS	4 (bases 1 to 170791)
TITLE	DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.
JOURNAL	Direct Submission
REFERENCE	Submitted (13-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
AUTHORS	On Mar 13, 2003 this sequence version replaced gi:15284265.
COMMENT	Draft Sequence Produced by DOE Joint Genome Institute
FEATURES	www.jgi.doe.gov
SOURCE	Finishing completed at Stanford Human Genome Center and Los Alamos National Laboratory
source	www.sbcg.stanford.edu
Location/Qualifiers	Quality: Phrap Quality >=40 100% of Sequence;
	Estimated Total Number of Errors is 0.
	1. 170791
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="16"
	/clone="RP11-147B17"

Oy		355	SerValGlyValGlyHisValAlaIaaArgGlyAlaIaaGAsnLysProProLeuGIYProAla	374
Dd		27829	AGGTGGGTGTGGGCCACCTGGCCCAAGGGCGAAGAACAAGCCCCCTCTGGGAACCGGC	27770
Oy		375	IleProAlaValSerProSerAlaHisIleAlaIaaserProAlaleuleuProSerIeu	394
Dd		27769	ATCCCTGGCGGTGTCCCCCTTCGGCCACCTGGCTGACGCCGCGCCTCTCCCTCCCTTA	27710
Oy		395	AlaProLeuGIYHisIleValSylSylSylHisIleAlaGAlaIalysGluSerGlnGlnIlyCyArg	414
Dd		27709	GCCCCCTCGGGCAAGAAGCAAGAACCGAGCCAAAGAAAGCACAGGGCTGGCCGG	27650
Oy		415	GlyLeuGlnAlaProleuAlaSerGlyGlyPProValIleuGIYARGLuHisleuArgGlu	434
Dd		27849	GGCTTCAGGACCACACTGGCTTACGTGGCCCTCTCTCTGGGGCGGGAGACCTTCGGCAG	27590
Oy		435	LeuProAlaleuValValTYrgIuSerGlnAlaGlyGlnProValGlnArgHisGluHis	454
Dd		27589	CTGGCCGCTTGTTGGTGTATGATGAGCCAGCCGGGCAACCGGTCCAGAGCATGAGCAC	27530
Oy		455	HisHisHisHisGluHisHisHisHisSTYrHisIlePheTYGlnThr	470
Dd		27529	CACACACACCATGACATCACACCATTCACCACTCTTACCAGACA	27482
RESULT 10				
LOCUS	AC145238			
DEFINITION	Pan troglodytes clone CH251-568P19, WORKING DRAFT SEQUENCE, 25 ordered pieces.	178790 bp	DNA	linear
ACCESSION	AC145238			
VERSION	AC145238.2			
KEYWORDS	HTGS; HTGS PHASE2; HTGS DRAPR.			
SOURCE	Pan troglodytes (chimpanzee)			
ORGANISM	Pan troglodytes			
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominae; Pan. 1 (bases 1 to 178790)			
AUTHORS	Antonellis,A., Aylee,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakeley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Cariga,K., Chu,G., Coleman,B., Coleman,H., Engle,J., Grantie,S., Guan,X., Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurtle,B., Idol,J.R., Karlins,E., Kwong,P., Latic,P., Lee-Lin,S.-Q., Legaia,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masetto,L.C., Maskeri,B., McQuell,J.J., Peguigner,C., Pearson,R., Portnoy,M.E., Prasad,A., Reddy-Dugue,N., Schandler,K., Schneider,M.G., Shah,K., Sison,C., Stanitipoo,S., Thomas,J.W., Thomas,P.D., Tsipouri,V., Vogt,J.L., Wetherby,K.D., Young,A. and Green,E.D.			
TITLE	NISC Comparative Sequencing Initiative			
JOURNAL REFERENCE	Unpublished 2 (bases 1 to 178790)			
AUTHORS	Green,E.D.			
TITLE	Direct Submission			
JOURNAL	Submitted (18-JUN-2003) NIH Intramural Sequencing Center, 8717 Grovemont Circle, Gaithersburg, MD 20877, USA			
REFERENCE	3 (bases 1 to 178790)			
AUTHORS	Green,E.D.			
TITLE	Direct Submission			
JOURNAL	Submitted (15-AUG-2003) NIH Intramural Sequencing Center, 8717 Grovemont Circle, Gaithersburg, MD 20877, USA			
COMMENT	On Aug 15, 2003 this sequence version replaced gi:31880081.			

The sequence data in this record represents an 'enhanced' version of a phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been

established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics -----

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171494 bases at least Q40
Consensus quality: 174120 bases at least Q30
Consensus quality: 175495 bases at least Q20
Insert size: 149000; agarose-fp
Quality coverage: 10.92x in Q20 bases; agarose-fp
Quality coverage: 9.22x in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently consists of 25 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 2604: contig of 2604 bp in length
2605 2704: gap of unknown length
2705 2790: contig of 5086 bp in length
7791 7890: gap of unknown length
7891 12833: contig of 4943 bp in length
12834 12933: gap of unknown length
12934 14664: contig of 1731 bp in length
14665 14764: gap of unknown length
14765 32237: contig of 17473 bp in length
32238 32337: gap of unknown length
32338 35356: contig of 3019 bp in length
35357 35456: gap of unknown length
35457 39292: contig of 3836 bp in length
39293 39392: gap of unknown length
39393 41210: contig of 1818 bp in length
41211 41310: gap of unknown length
41311 43762: contig of 2452 bp in length
43763 43862: gap of unknown length
43863 46766: contig of 2904 bp in length
46767 46866: gap of unknown length
46867 53405: contig of 6539 bp in length
53406 53505: gap of unknown length
53506 55334: contig of 1829 bp in length
55335 55435: gap of unknown length
55436 59095: contig of 13661 bp in length
59096 69195: gap of unknown length
69196 78811: contig of 9616 bp in length
78812 78911: gap of unknown length
78912 89715: contig of 10804 bp in length
89716 89815: gap of unknown length
89816 97259: contig of 7444 bp in length
97260 97359: gap of unknown length
97360 99238: contig of 1879 bp in length
99239 99338: gap of unknown length
99339 101324: contig of 1986 bp in length
101325 101424: gap of unknown length
101425 112068: contig of 10644 bp in length
112069 112168: gap of unknown length
112169 118679: contig of 6511 bp in length
118680 118779: gap of unknown length
118780 120680: contig of 1901 bp in length
120681 120780: gap of unknown length
120781 135732: contig of 14952 bp in length

FEATURES

* 135733 135832: gap of unknown length
* 135833 138723: contig of 2891 bp in length
* 138724 138823: gap of unknown length
* 138824 163652: contig of 24829 bp in length
* 163653 163752: gap of unknown length
* 163753 178790: contig of 15038 bp in length.
Location/Qualifiers

Source

1. 178790
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone_1b="CH251-568P19"
/clone_1b="CH251"
1. 90615
/note="clone overlaps with GenBank Accession Number AC145175 clone CH251-32B1 (center project name esj)"
1. 2604
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
2705. 7790
/note="assembly_fragment"
7891. 12833
/note="assembly_fragment"
12934. 14664
/note="assembly_fragment"
14765. 32237
/note="assembly_fragment"
32338. 35356
/note="assembly_fragment"
35457. 39292
/note="assembly_fragment"
39393. 41210
/note="assembly_fragment"
41311. 43762
/note="assembly_fragment"
43863. 46766
/note="assembly_fragment"
46867. 53405
/note="assembly_fragment"
53506. 55334
/note="assembly_fragment"
55435. 59095
/note="assembly_fragment"
59196. 78811
/note="assembly_fragment"
78912. 89715
/note="assembly_fragment"
89816. 97259
/note="assembly_fragment"
97360. 99238
/note="assembly_fragment"
99339. 101324
/note="assembly_fragment"
101425. 112068
/note="assembly_fragment"
112169. 118679
/note="assembly_fragment"
118780. 120680
/note="assembly_fragment"
120781. 135732
/note="assembly_fragment"
135824. 163652
/note="assembly_fragment"
163753. 178790
/note="assembly_fragment"
clone_end:T7
vector_side:right"


```
* 152074 160497: contig of 8424 bp in length
* 160498 160597: gap of unknown length
* 160598 170482: contig of 9885 bp in length
* 170483 170582: gap of unknown length
* 170583 207486: contig of 36904 bp in length.
FEATURES
    source
        1. .207486
            /organism="Pan troglodytes"
            /mol_type="genomic DNA"
            /db_xref="taxon:9598"
            /clone="CH251-3281"
            /clone_11b="CH251"
            1. .430
            /note="assembly_fragment"
            clone_end:SP6
            vector_bide:left"
            misc_feature
                /note="assembly_fragment"
                531..20321
            misc_feature
                /note="assembly_fragment"
                20422..99662
            misc_feature
                /note="assembly_fragment"
                96974..207486
            /note="clone overlaps with GenBank Accession Number
            AC145238 clone CH251-568P19 (center project name esk)"
            misc_feature
                /note="assembly_fragment"
                99763..106538
            misc_feature
                /note="assembly_fragment"
                106639..113489
            misc_feature
                /note="assembly_fragment"
                113590..143460
            misc_feature
                /note="assembly_fragment"
                143561..149849
            misc_feature
                /note="assembly_fragment"
                149950..151973
            misc_feature
                /note="assembly_fragment"
                152074..160497
            misc_feature
                /note="assembly_fragment"
                160598..170482
            misc_feature
                /note="assembly_fragment"
                170583..207486
            /note="assembly_fragment
            missing 17 clone end on 3' end of insert"

ORIGIN
Alignment Scores:
Pred. No.:      4.19e-77      Length:      207486
Score:          88.00         Matches:    194
Percent Similarity: 97.98%      Conservative: 0
Best Local Similarity: 97.98%      Mismatches: 2
Query Match:    18.72%         Indels:     4
DB:             2            Gaps:        0

US-09-993-966-7 (1-470) x AC145175 (1-207486)
QY 275 GlycerProSeValAlaGlnLySeSerGluLeuProProArgThrSerAnProThrArg 294
Db 180467 GGTCCTCCCTTCCTGCGCCGAGAGTCAAGACTCCCCCGGCACCTCCATCCACCTCGA 180526
QY 295 SerArgSerHisGluProGluAlaIleHisIle-ProHisArgLySerProGlnGlyValAs 314
Db 180527 TCTCGCTCCCATGAGCCGGAAGCATTCACGT-CCGACACCGAAAGCCCAAGCGCGTGA 180585
QY 314 pProAlaSerPheHisPheLeuSerPheProIleAlaValSerGluLeuGlnGlnAT 334
Db 180586 CCGGCGCTCTCCATCTCTTACACACCCCAATCGCCAAAGTCTCAGAGCTCCAGCAACG 180645
QY 334 gLeuATGGLyThGlnAlaPbGlySerLyHisPheValAlaArgSerProLyAlaGlnGlyLy 354
Db 180646 GCTCCGGGGGACCCAGAGCGGAGCAAGCACTTTGTGAGSTCCGCCCAAGGCCCAAGGGCAA 180705
QY 354 sSerValGlyValGlyHisValAlaArgGlyAlaArgGlyAlaArgGlyAlaArgGlyAla 374
Db 180706 GAGTGTGGGTGTGGGCACTGTGCGCCAGAGGGGCAAGAAACAAAGCCCTCTGGGACCGGC 180765
QY 374 alleProAlaValSerProSerAla-HisLeuAlaSerProAlaLeuLeuProSerL 394
```

```
Db 180766 CATCCCTGCGGTGTCCCTCCAC-CCACCTGCTGCCAGCCGCTCTCCCTCC 180824
QY 394 euaIaProLeuGlyHisLySerLyHisLySerLyHisLySerLyHisLySerLyHisLySer 414
Db 180825 TAGCCCCCTCGGGACCAAGAACACACAGCCGAGCCAGGAGGAGGAGGAGGAGGAGGAGG 180884
QY 414 rGGLyLeuGlnAlaProLeuAlaSerGlyGlyProValLeuGlyAlaGlyHisLeuArg 434
Db 180885 GGGGCGCTGACGAGCACCATGCGCTCAGGTGCGCTGTCTGGGGCGGAGCACCCTGCGGG 180944
QY 434 lueuProAlaLeuValAlaValTyGlySerGlnAlaGlyGlnProValGlnArgHisGly 454
Db 180945 AGCTGCGCGCTGTGTGTGTACAGAGCCAGCGCGGAGCGGCTCCAGACATAGC 181004
QY 454 ishHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 470
Db 181005 ACCACCAACACATGAACATCACACATTAACACACATCTTACAGACA 181054

RESULT 12
AC145003
LOCUS
DEFINITION
AC145003 194869 bp DNA linear HTG 03-JUN-2003
Papio anubis clone RP41-205A15, WORKING DRAFT SEQUENCE, 9 ordered
pieces.
AC145003
VERSION
AC145003.2 GI:32441289
HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE
Papio anubis
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
Cercopithecinae; Papio.
1 (bases 1 to 194869)
REFERENCE
AUTHORS
Antonellis,A., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Cariaga,K.,
Chu,G., Coleman,B., Coleman,H., Engle,J., Granite,S., Guan,X.,
Gupte,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,
Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Latic,P., Lee-Lin,S.-Q.,
Legaspi,R., Machado,Q.L., Machado,V.B., Marquilles,E.H., Mastello,C.,
Maskeri,B., McDowell,J., Paguirigan,C., Pearson,R., Portnoy,M.E.,
Prasad,A., Reddix-Duque,N., Schandler,K., Schueler,M.G., Shah,K.,
Sison,C., Stantropop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V.,
Vogt,J.L., Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 194869)
REFERENCE
AUTHORS
Green,E.D.
TITLE
Direct Submission
JOURNAL
Submitted (30-MAY-2003) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 194869)
REFERENCE
AUTHORS
Green,E.D.
TITLE
Direct Submission
JOURNAL
Submitted (03-JUN-2003) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA
On Jul 3, 2003 this sequence version replaced gi:31193878.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
Center project name: dtr
Center clone name: 205A15

The sequence data in this record represents an 'enhanced'
version of a Phase 2 submission. Specifically, the indicated
order and orientation of each sequence contig has been
established using one or more of the following: read-pair
data from individual subclones, overlaps with neighboring
clones, alignment with available reference sequence (e.g.,
human), and/or confirmation by PCR testing. In addition,
the sequence assembly is based on at least 8x average
```

coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 193209 bases at least Q40
 Consensus quality: 193760 bases at least Q30
 Consensus quality: 193984 bases at least Q20
 Insert size: 180000; agarose-fp
 Quality coverage: 14.34x in Q20 bases; agarose-fp
 Quality coverage: 13.30x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and the accession number will be preserved.

```

1       7608: contig of 7608 bp in length
*       7609       7709: gap of unknown length
*       7709       10151: contig of 2443 bp in length
*       10152       10251: gap of unknown length
*       10252       44876: contig of 34625 bp in length
*       44877       44977: gap of unknown length
*       44977       76344: contig of 31368 bp in length
*       76345       76444: gap of unknown length
*       76445       78447: contig of 2103 bp in length
*       78448       78647: gap of unknown length
*       78648       145819: contig of 67172 bp in length
*       145820       163253: contig of 17334 bp in length
*       163254       163353: gap of unknown length
*       163354       166118: contig of 2765 bp in length
*       166119       166219: gap of unknown length
*       166219       194869: contig of 28651 bp in length.

```

FEATURES

source

```

1. 194869
   /organism="Papio anubis"
   /mol_type="genomic DNA"
   /db_xref="taxon:9555"
   /clone="RP41-205A15"
   /clone_1b="RP41"

1. 31748
   /note="clone overlaps with GenBank Accession Number
AC145043 clone RP41-2E13 (center project name dtg)"

misc_feature
   /note="assembly_fragment
clone end:T7
vector side:left"

7709..10151
   /note="assembly_fragment"

10252..44876
   /note="assembly_fragment"

44977..76344
   /note="assembly_fragment"

76445..78547
   /note="assembly_fragment"

78648..145819
   /note="assembly_fragment"

145920..163253
   /note="assembly_fragment"

163354..166118
   /note="assembly_fragment"

166219..194869
   /note="assembly_fragment
clone end:SP6

```

ORIGIN

vector_side:right"

Alignment Scores: 1.32e-46 Length: 194869
 Pred. No.: 58.00 Matches: 108
 Score: 98.18% Conservative: 0
 Percent Similarity: 98.18% Mismatches: 1
 Best Local Similarity: 12.34% Indels: 2
 Query Match: 2 Gaps: 0
 DB: 2

US-09-993-966-7 (1-470) x AC145003 (1-194869)

```

Qy      317 SerPheHisPheLeuAapThrProIleAlaIysValSerGIuEugIngInIArgLeuArg 336
Db      16652 TCCTTCACACTCTCTTGACACCCCAATCGCAAGCTCTCAAGCTCCAGCAAGCGCTCCG 16711
Qy      337 G1YThrGlnAapGIyserylshIsPheValArgSerProIySaIagInGIyServal 356
Db      16712 GGCACCCAGCAGCGGAGCAAGCACTTGTGAGGTCCCAAGGCCCGGCAAGAGCGGTG 16771
Qy      357 G1YValG1YHisValAlaAArgG1YAlaArgAsnIyProIeuGIyProAlaIle-Pr 376
Db      16772 GGTGTGGGCCACGTCGGCCAGAGGGCGAAGAAAGCCCTCTCGGACCCGCTT-CCC 16830
Qy      376 oAlaValSerProSerAlaHisLeuAlaAIsaSerProAlaLeuEupProSerLeuAlaPr 396
Db      16831 TGCAGTCTCCCCCTCCGCCCACTGGCGCCGCAAGCCCGCCCTCTCCCTCCTTAGCCCC 16890
Qy      396 oLeuGIYHisIyValshIsIyshIsArgAlaIyGIuSerGIuIngInIyCyArgGIyLe 416
Db      16891 CTCGGGCAACAAGAGCAAGCAAGCCAGCAAGAGAGCAGCAGGAGGTGCGGGGCGCT 16950
Qy      416 uGlnAlaProLeuAlaSerGIyGIyPro 425
Db      16951 ACAAGCACCGCTGCGCTCGGTGCGCC 16978

```

RESULT 13

LOCUS

AC145043 199189 bp DNA linear HTG 26-JUN-2003
 DEFINITION Papio anubis clone RP41-2E13, WORKING DRAFT SEQUENCE, 4 ordered pieces.

AC145043
 VERSION AC145043.2 GI:32261339
 KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT.
 SOURCE Papio anubis (olive baboon)
 ORGANISM Papio anubis

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecoidea; Papio.
 1 (bases 1 to 199189)

REFERENCE
 AUTHORS Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakeley,R.W., Boufford,G.G., Brinkley,C., Brooks,S., Cariaga,K., Chu,G., Coleman,B., Coleman,H., Engle,J., Granite,S., Guan,X., Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-D., Hu,P., Hurle,B., Idol,J.R., Karins,B., Kwong,P., Laric,P., Lee-Edlin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B., Marquies,E.H., Masello,C., Maskeri,B., McDowell,U., Paguirigan,C., Pearson,R., Portnov,M.E., Prasad,A., Reddy-Dugue,N., Schander,K., Schueler,M.G., Shah,K., Sison,C., Stantirpop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V., Vogt,J.L., Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
 NISC Comparative Sequencing Initiative

TITLE JOURNAL
 REFERENCE
 AUTHORS Green,E.D.
 TITLE Direct Submission
 JOURNAL Submitted (04-JUN-2003) NIH Intramural Sequencing Center, 8717
 Groveomont Circle, Gaithersburg, MD 20877, USA
 3 (bases 1 to 199189)

REFERENCE
 AUTHORS Green,E.D.
 TITLE Direct Submission
 JOURNAL Submitted (26-JUN-2003) NIH Intramural Sequencing Center, 8717
 Groveomont Circle, Gaithersburg, MD 20877, USA

COMMENT

On Jun 26, 2003 this sequence version replaced gi:31376430.

----- Genome Center -----
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc200@hgrl.nih.gov
Project Information
Center project name: dcg
Center clone name: 002c3

The sequence data in this record represents an 'enhanced' version of a phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones; overlaps with neighboring clones; alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

```

Summary statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 197710 bases at least Q40
Consensus quality: 197813 bases at least Q40
Consensus quality: 197876 bases at least Q20
Insert size: 197000; agarose-fp
Insert size: 197889; sum-of-configs
Quality coverage: 11.54x in Q20 bases; agarose-fp
Quality coverage: 11.54x in Q20 bases; sum-of-configs

```

NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submittor.

This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

1	20663:	contig of 20663 bp in length	
*	20664	20763:	gap of unknown length
*	20764	173968:	contig of 153225 bp in length
*	173989	174088:	gap of unknown length
*	174089	176651:	contig of 2563 bp in length
*	176652	176751:	gap of unknown length
*	176752	198189:	contig of 21438 bp in length.

FEATURES	Location/Qualifiers
source	1. .198189

```
misc_feature      1..20663
                  /note="assembly_fragment
clone_end:SP6
vector_side:left"
misc_feature      20764..17398
                  /note="assembly_fragment"
misc_feature      166409..198189
                  /note="clone overlaps with GenBank accession Number
AC145003 clone RP41-205A15 (center project name drr)"
misc_feature      174089..116651
                  /note="assembly_fragment"
misc_feature      176752..198189
```

ORIGIN

Alignment Scores:	
Pred. No.:	1,34e-46
Score:	58.00
Percent Similarity:	98.18
Best Local Similarity:	98.18
Query Match:	12.34
DB:	2
Length:	18189
Matches:	108
Conservative:	0
Mismatches:	1
Indels:	2
Gaps:	0

US-09-993-966-7 (1-470) X AC145043 (1-198189)

Qy	Db	Qy
317	183092	337
serPheHisPheLeuAspThrProIleAlaValSerGluLeuGlnGlnArgLeuArg	TCTTTCACCTTCCTTGACACCCCAATGCCCAAGCTCTGAGACTCCAGCAAGGCTCCGG	GLYThrGlnArapGlySerLeuHisHisPheValArgSerProGlyAlaGlnGlyLeuSerVal
183152	183151	183152
GGCACCCACGAGCGGAGACACACTTTGTGAGAGTCCCCCAAGGCCCAAGGCAAGACGCG		

[illegible]

QY	396	OLEUGL ^{YH} SL ^Y LYSL ^{YSH} SL ^Y YSH ^{SL} YAR ^{YAL} AL ^Y SG ^Y LS ^Y SR ^Y GL ^Y NG ^Y ING ^Y LY ^Y CYS ^Y AR ^Y GL ^Y YLE	416
Db	183331	CCTCGGCGCAAGACGACACGACCGACCGACGAGACGCCACGACGGCGCTCCCGGGGCGCT	183399

```

QY      416 uGlnAlaProLeuAlaSerGlyGlyPro 425
          |||||
Db      183391 ACAGCACC GCTGGCTCGGGGTGCCCC 183418

```

RESULT 14	LOCUS	AX172351	DEFINITION
	AX172351	1401 bp	DNA
	Sequence 1 from Patent WO0144279.	1linear	PAT 03-JUL-2001

SOURCE	MUS	MUSCULUS	(house mouse)
ORGANISM	Mus	musculus	

REFERENCE	AUTHORS	TITLE	JOURNAL
1	Yan, D. and Williams, L. T.	Mammalian dishevelled-associated proteins	Patent: WO 0144279-A 1 21-JUN-2001; Chiron Corporation (US)

ORIGIN

Alignment Scores:	
Pred. No.:	2,18e-39
Score:	49.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	10.43%
BB:	6
Length:	1401
Matches:	49
Conserves:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-09-993-966-7 (1-470) x AX172351 (1-1401)

QY 131 G L U G I A A P S S e r A r g G I n G l u T r P h r e h m r e u T y r A p P h e a p a n A n g I y L y s 150
Db 391 G A G S A G A C A G C G C C A G A G T G A C T T C A C T A T A T A C T T C G A C A A C A A T G C C A A 450
QY 151 V a l T h r A r g I u A p l l e T h r S e r l e u e n i s h r l l e T y r G l u V a l V a l A p S e r S e r 170

Db 451 GTGACCGGTGAGGACATTACAGCTTGCTGCATCATCATATGAAAGTGTGACTCCTCT 510
 QY 171 ValAsnHisSerProThrSerSerIys 179
 Db 511 GTGAACCATTTCCCCACATCAAGCAAG 537

Db 511 GTGAACCATTTCCCCACATCAAGCAAG 537
 Search completed: December 30, 2004, 12:16:23
 Job time : 8674 secs

RESULT 15
 AF343352 1416 bp mRNA linear ROD 29-MAR-2001
 LOCUS AF343352
 DEFINITION Mus musculus Nkd mRNA, complete cds.
 ACCESSION AF343352
 VERSION AF343352.1 GI:13487304
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 TITLE 1 (bases 1 to 1416)
 JOURNAL Yan,D., Wallingford,J.B., Sun,T.Q., Nelson,A.M., Sakanaka,C., Reinhard,C., Harland,R.M., Pantl,W.J. and Williams,L.T.
 MEDLINE Cell autonomous regulation of multiple Dishevelled-dependent pathways by mammalian Nkd
 PUBMED Proc. Natl. Acad. Sci. U.S.A. 98 (7), 3802-3807 (2001)
 11274398

REFERENCE
 AUTHORS 2 (bases 1 to 1416)
 TITLE Yan,D., Wallingford,J.B., Sun,T.-Q., Nelson,A.M., Sakanaka,C., Reinhard,C., Harland,R.M., Pantl,W.J. and Williams,L.T.
 JOURNAL Direct Submission
 Submitted (26-JAN-2001) Research, Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608, USA
 FEATURES

source
 1. 1416
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 1. 1416
 /note="mNkd; similar to Drosophila naked cuticle"
 /codon_start=1
 /product="Nkd"
 /protein_id="AAK27485.1"
 /db_xref="GI:13487305"
 /translation="MGKLHSKPAVNCRRSPESGDSFPAVSAAARKIEEMIGRQCP
 GSVGPRQLAGTVGRTGLVGRTRBALGEDEDPLVAVLPPEKIDSLGSGE
 KRMERLSEPGQASPKQLKREELQCDVSEBDSQSKSVLFNPDNGKYTRREDITSL
 HTTVEVDSSVNSPTSKTLRVKLTVAPOGSKRSVLFNPDNGKYTRREDITSL
 ELRGWEKKQAPLRFQGDSDLBDPCVHNCVDENIERRNHYLIDLGIENYTSQPGPS
 PSVAOKSELPRISNPTSRSSHEPEAAHIDRRPOGVDPGSFHLDTPPAKASELQOR
 LRGTDPSKHFVRSPPKAGKMGWGAGARSKRPPLVPTTHTVSPSAHLATSPALLP
 TLALGHKKHKAKESSQASCRGLQGPLAAGSTVMGREQVRELPAVVVYESQAGAV
 QREHHNHHHHHHHHFYQP"

ORIGIN
 Alignment Scores:
 Pred. No.: 2.21e-39 Length: 1416
 Score: 49.00 Matches: 49
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 10.43% Indels: 0
 DB: 10 Gaps: 0

US-09-993-966-7 (1-470) x AF343352 (1-1416)

QY 131 GUGUUAapSerAaGInGUtTPTThPheThIeuTyraapPheAapAaenAangIyLys 150
 Db 391 GAGGAGAGACAGCCGCGCAAGTGACATTTCCTATATGACTTCGACAAATGGCAAA 450
 QY 151 ValThrArgGUaapIleThSerIeuIenHisThrIleTyrgIuValaAapSerSer 170
 Db 451 GTGACCGGTGAGGACATTACAGCTTGCTGCATCATCATATGAAAGTGTGACTCCTCT 510
 QY 171 ValAsnHisSerProThrSerSerIys 179